Optimal Experimental Design for Nonlinear and Generalised Linear Models

By

Timothy Hugh Waterhouse, B.Sc. (Hons)

A thesis submitted to the University of Queensland

for the degree of Doctor of Philosophy

School of Physical Sciences

September 2005

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Parts of Chapter 3 have been published in Waterhouse *et al.* (2004a) and parts of Chapter 4 have been accepted for publication in Waterhouse *et al.* (2005).

Timothy Hugh Waterhouse, B.Sc. (Hons)

Acknowledgements

This thesis would not have been possible without the expert guidance of my PhD advisors, Prof John Eccleston and Dr Stephen Duffull, whose contagious enthusiasm for the topic has kept me motivated and inspired throughout my degree. I am grateful to have been able to benefit from their combined experience. They have provided ample opportunity to attend conferences, both statistical and pharmacological, where I was able to meet and share ideas with many like-minded people.

James McGree assisted with the creation of the MATLAB code for the cross-entropy method, which also benefited from the help of Dr Dirk Kroese.

My 2004 internship with Michael Heathman and others at Eli Lilly and Co. in Indianapolis provided a unique opportunity to apply knowledge gained during this degree to real-world problems, giving me an improved understanding of optimal design in practice.

My friends, family and fellow PhD students have made the whole degree more enjoyable, providing the necessary support and distraction when needed. Special thanks must go to the unyielding metaphoric whip-cracking of certain individuals during the last few months of writing.

This thesis is dedicated to my mother.

Abstract

Much of the literature on optimal design of experiments has focussed on experiments where the behaviour of the system is approximated by a linear model, such as a low-order polynomial. In many areas such as pharmacology and chemistry, such approximations are not appropriate, as the underlying mechanisms produce highly nonlinear or categorical responses. This thesis addresses some issues with the optimal design of experiments in these situations.

Commonly used criteria for the 'optimal' design of experiments relate to optimality in terms of efficient estimation of model parameters. However, quite often another important objective of an experiment is to select the model structure which best describes the underlying behaviour of the system. We examine existing criteria for model discrimination for both nonlinear and generalised linear models, and combine them with criteria for parameter estimation in order to create designs which address both objectives. We show that these designs can be quite efficient in terms of each of the criteria.

Further complications in the design process arise with the use of mixed effects models, that is when some model parameters are allowed to vary randomly between blocks or clusters of units. The Fisher information matrix is involved in the calculation of many optimality criteria, and this matrix cannot be written down in closed form for many nonlinear and generalised linear mixed effects models. We instead rely on approximations to the information matrix to generate optimal designs. This thesis gives details of the use of an existing approximation to the matrix in the optimal design of a complex pharmacokinetic experiment involving nonlinear mixed effects models. We also investigate several alternative approximations to the matrix for logistic regression with random coefficients, with an application in pharmacodynamics: the design of a cross-over trial with a binary response.

Regardless of the criterion used to select a particular design, we require a method to search the design space for points which maximise or minimise the criterion. The models considered in this thesis are assumed to have predictor variables taken over a continuous range, so combinatorial optimisation techniques such as the tabu algorithm are not appropriate. Instead we make use of the simulated annealing algorithm (modified for continuous variables) and a relatively new algorithm known as the cross-entropy method. Both algorithms are implemented in programs written for the MATLAB package.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The increasing use of optimisation techniques and modern powerful computers has lead to the use of more complex and realistic models in data modelling. However, judging by a review of the literature in the optimal design of experiments (Silvey, 1980; Atkinson and Donev, 1992; Pukelsheim, 1993), the majority of research in this area remains focussed on linear models, while in fact nonlinear and generalised linear models (GLMs) are commonly used in many areas such as chemistry and pharmacology. Atkinson and Donev (1992), for example, gives an excellent introduction to the theory of optimal design of experiments, but only devotes one chapter to nonlinear models, and a section of the 'Further Topics' chapter on generalised linear models.

Optimal design has been studied since as early as 1918, when the paper by Smith (1918) defined the objective of minimising the worst-case prediction error in the construction of polynomial models. In this paper, she introduced what is now known as G-optimality (which minimises the maximum over the design space of the standardised variance, as defined by Kiefer and Wolfowitz (1959)). In the same paper by Kiefer and Wolfowitz, they gave the name of D-optimality to the criterion introduced by Wald (1943), which puts the emphasis on the quality of the parameter estimates. A D-optimal design maximises the determinant of the Fisher information matrix, hence minimising the volume of the confidence ellipsoid of any unbiased parameter estimates. Kiefer and Wolfowitz (1959) also relate these two alphabetic criteria by the General Equivalence theorem.

All of the papers mentioned so far only consider models which are linear in their parameters, i.e. models of the form

$$\boldsymbol{y} = \eta(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{1.1}$$

for explanatory variables $\boldsymbol{X}$, model parameters $\boldsymbol{\theta}$, and zero-mean residual error terms $\boldsymbol{\epsilon}$ (these models are introduced more formally in the next chapter). Certain complications arise in the optimal design of experiments when the function η is not linear in the parameters, the most notable of which is that nonlinear models require estimates of the model parameters to be known for an optimal design to be constructed. Typically initial parameter estimates are based on the results of previous studies or 'expert' guesses. Optimal designs found using point estimates are said to be locally optimal.

Although work on the theory of optimal design for linear models dates back as far as 1918, nonlinear models did not appear in the optimal design literature until 1959. Box and Lucas (1959) investigated locally D-optimal designs for nonlinear models, in which the authors justified the dependence of such designs on prior knowledge of parameter values by stating that "in practical problems it will almost invariably be the case that some such information is available, and this will then provide the basis of a first design". The literature on optimal design for nonlinear models has been relatively sparse due to this dependence on parameter values. (One approach to avoiding the need for point estimates in optimal design is the use of Bayesian experimental designs, in which prior distributions are placed on the model parameters. These Bayesian approaches are outside the scope of this thesis, but are reviewed in Chaloner and Verdinelli (1995) and Clyde (2001).)

In the area of pharmacokinetics (PK, the study of the time course of the changing concentration of a drug in the body), for example, where nonlinear models are commonly used, most work on optimal design has been theoretical, with little in the way of application. The literature in the area begins with D'Argenio (1981), who discussed the optimal choice of sampling times in a PK study, but confined the optimisation to a single subject.

The increasing use of mixed effects models (such as random coefficient models and models with random block effects) has led to a growing interest in optimal design for such models. Optimal design for linear mixed effects models has been addressed by several papers (Cheng,

1995; Atkins and Cheng, 1999; Berger and Tan, 2004). Mentré *et al.* (1997) propose an approach to optimal design for random-effects regression models in which nonlinear models are first linearised. This paper by Mentré *et al.* has led to an increasing use of optimal design in population PK studies (i.e. where the models involve between subject variability). Two prospective evaluations of an optimally designed PK study have been published (Mentré *et al.*, 2001; Green and Duffull, 2003), with the third (Waterhouse *et al.*, 2005) recently submitted for publication, also appearing as a chapter of this thesis.

There have been several papers on optimal design for efficient estimation of parameters in generalised linear models (Chaloner and Larntz, 1989; Atkinson *et al.*, 1995; Atkinson and Haines, 1996), but the literature in this area tends to be even more sparse than in the area of nonlinear models, due to the increased theoretical and computational challenges associated with GLMs. GLMs are commonly used in experiments in areas such as chemistry, pharmacology and engineering (see Myers *et al.* (2002) for examples of applications), so optimal design for these models warrants further consideration.

Longford (1994) is concerned with estimation in logistic regression models with random coefficients, and presents an approximation to the Fisher information matrix which may be useful in optimal design. However, there appears to be very little, if any, literature on the topic of optimal design for generalised linear mixed models. This thesis addresses some formidable problems associated with the construction of optimal designs for these models.

The majority of work on optimal design for nonlinear and generalised linear models has focussed on parameter estimation. Another common goal of an experiment is to choose between two or more competing model structures. Atkinson and Fedorov (1975a) introduced the T-optimality criterion for discriminating between two nonlinear models, in which one of the two models is assumed to be true (with known parameter values) and the noncentrality parameter for lack of fit of the false model is maximised. This was generalised to accommodate for more than two competing models in Atkinson and Fedorov (1975b), and analogous criteria were defined for generalised linear models in Ponce de Leon and Atkinson (1992).

This thesis addresses a number of issues related to optimal design for nonlinear and generalised linear models. After introducing some notation and basic concepts of optimal design (including the search methods used in this thesis) in Chapter 2, the thesis is divided

into two parts: Part I relates to nonlinear models, while Part II is concerned with generalised linear models.

Chapter 3, the first chapter in Part I, describes a number of methods for finding designs which are near-optimal in terms of both parameter estimation and model discrimination. Several examples are given to demonstrate their effectiveness. Chapter 4 describes the construction of an optimal design involving nonlinear mixed effects models with multiple responses, where the objective of the design is to discriminate between two competing models as well as estimate the parameters well under each model. The work is motivated by a real-life pharmacokinetic study.

Part II starts with Chapter 5, an analogy to Chapter 3 for generalised linear models, in which the aim is to construct a design which is useful in terms of both parameter estimation and model discrimination. The effectiveness of the designs in terms of model discrimination is evaluated by power tests. Chapter 6 is motivated by another example in pharmacometrics, a pharmacodynamic study of a drug with a binary response. A range of locally optimal parallel (single period) and crossover designs are constructed for a logistic regression model, and the sensitivities of the designs to misspecification of the parameter values are examined. A union of parallel and crossover designs is also considered, and the loss of efficiency resulting from allocating more patients to the parallel group is investigated. The last chapter in Part II, Chapter 7, outlines a number of approaches to optimal design for generalised linear mixed models. The model used in the example of Chapter 6 is extended to allow the parameters to vary randomly between patients and the corresponding optimal designs are compared to the designs for the fixed effects model in the previous chapter.

The thesis ends with some concluding remarks and discussion of future work in the area in Chapter 8.

Chapter 2

Preliminaries

This chapter introduces the concept of an experimental design, the types of models that are used in this thesis, as well as some criteria used to optimise the designs for these models, for a number of objectives. Finally, the search algorithms used to perform these optimisations are described.

2.1 Approximate and exact experimental designs

An experimental design ξ with n support points can be written as

$$\xi = \begin{Bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix}, \tag{2.1}$$

where $\boldsymbol{\xi}_i \in \mathcal{X}$ are the support points (m-vectors also known as elementary designs) consisting of the explanatory variables which describe the experimental conditions, with weights $w_i \in [0, 1]$ summing to 1. The design space can be written $\Xi = \{\xi \in \mathcal{X}^n \times [0, 1]^n : \sum_{i=1}^n w_i = 1\}$. For an exact design with N experimental units, the weights are given explicitly by $w_i = r_i/N$, where r_i is the number of replications of $\boldsymbol{\xi}_i$. For an approximate (or 'continuous') design, the weights also represent the proportion of experimental effort at each point, but the design is not constrained to apply to a fixed number of experimental units. Methods for choosing this design in an 'optimal' way are described below.

2.2 Models

2.2.1 Linear models

In the context of this thesis, *linear models* are considered to be models which are linear in their parameters, that is they are of the form

$$\boldsymbol{y} = \eta(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{2.2}$$

for the $n \times p$ matrix $\boldsymbol{X} = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_n')'$ (the elements of $\boldsymbol{x}_i$ are functions of the support points $\boldsymbol{\xi}_i$) and the $p \times 1$ vector of parameters $\boldsymbol{\theta}$, where the elements of $\boldsymbol{\epsilon}$ are the residual error terms, usually assumed to be independently normally distributed with zero mean and constant variance σ^2, and '$'$' denotes the vector/matrix transpose. A common example of a linear model is a polynomial such as the quadratic with one explanatory variable:

$$\eta(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2.$$

These types of models have been covered extensively in the optimal design literature, and are not the focus of this thesis.

2.2.2 Nonlinear models

The use of *nonlinear models* gives rise to some interesting problems in optimal design. The models are considered to be of the form

$$\boldsymbol{y} = \eta(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}, \tag{2.3}$$

for the $n \times p$ matrix $\boldsymbol{X}$ and the $p \times 1$ vector of parameters $\boldsymbol{\theta}$, where the elements of $\boldsymbol{\epsilon}$ are again typically assumed to be independently normally distributed with zero mean and constant variance σ^2. This is a generalisation of the model in Equation (2.2), where the function η is no longer constrained to be linear in the model parameters, that is it cannot be expressed as $\boldsymbol{X}\boldsymbol{\theta}$. One example of a nonlinear model is the three-parameter quadratic Michaelis-Menten model with a single explanatory variable:

$$\eta(x, \boldsymbol{\theta}) = \frac{\theta_1 x}{\theta_2 + x + \theta_3 x^2}.$$

2.2.3 Generalised linear models

Generalised linear models (GLMs) encompass a large class of models, from simple linear regression models to models for quantal responses to models for survival data. They can be studied as a single class, and are all defined by three characteristics:

1. The distribution of the $n \times 1$ vector of independent responses, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, with means $\mathrm{E}(Y_i) = \pi_i$ and variances $\mathrm{Var}(Y_i) = a(\phi)V(\pi_i)$, where $a(\phi)$ is a scale factor which doesn't depend on π_i. We write $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)'$.

2. The linear predictor

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)' = \boldsymbol{X}\boldsymbol{\theta},$$

   where $\boldsymbol{X}$ is the $n \times p$ matrix of known functions of the m explanatory variables, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ is the $p \times 1$ vector of model parameters.

3. The link function g, providing the link between the mean vector $\boldsymbol{\pi}$ and the linear predictor $\boldsymbol{\eta}$:

$$g(\pi_i) = \eta_i.$$

The GLMs considered in this thesis are generally for binomial data, where S_i successes are observed from m_i trials (i.e. $S_i \sim \mathrm{Bin}(m_i, \pi_i)$), and the proportion of successes is written as $Y_i = S_i/m_i$. The expected value of Y_i is then the probability of success π_i. Some common link functions for binomial data are the logit link function, $g(\pi_i) = \mathrm{logit}(\pi_i) = \log\{\pi_i/(1-\pi_i)\}$; the probit link function, $g(\pi_i) = \Phi^{-1}(\pi_i)$, where Φ is the cumulative distribution function of the standard normal distribution; and the complementary log-log link function, $g(\pi_i) = \log\{-\log(1-\pi_i)\}$.

It is also worth defining here the deviance of the fit of a generalised linear model for binomial data:

$$\begin{aligned} D(\boldsymbol{Y}; \hat{\boldsymbol{\pi}}) &= 2\ell(\boldsymbol{Y}; \boldsymbol{Y}) - 2\ell(\hat{\boldsymbol{\pi}}; \boldsymbol{Y}) \\ &= 2\sum_{i=1}^{n} w_i \left[Y_i \log(Y_i/\hat{\pi}_i) + (1-Y_i)\log\left(\frac{1-Y_i}{1-\hat{\pi}_i}\right)\right], \end{aligned} \tag{2.4}$$

where $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \ldots, \hat{\pi}_n)'$ is the vector of fitted values of $\boldsymbol{\pi}$ from the data $\boldsymbol{Y}$, and ℓ is the log-likelihood. For observed values $\boldsymbol{y} = (y_1, \ldots, y_n)'$ of the random variables in $\boldsymbol{Y}$, the log-likelihood can be written

$$\ell(\boldsymbol{\pi}; \boldsymbol{y}) = \sum_{i=1}^{n} \left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + m_i \log(1-\pi_i) \right].$$

2.3 Design criteria

2.3.1 Fisher information matrix

For a model involving the parameter vector $\boldsymbol{\theta}$ and an experimental design ξ as described in Equation (2.1), the expected Fisher information matrix is defined as

$$\boldsymbol{M}(\boldsymbol{\theta}, \xi) = \mathrm{E}\left[-\frac{\partial^2 \ell(\boldsymbol{\theta}; \boldsymbol{Y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right],$$

where $\ell(\boldsymbol{\theta}; \boldsymbol{Y})$ is the log-likelihood of the $N \times 1$ vector of observations $\boldsymbol{Y} = (Y_1, \ldots, Y_N)'$.

For linear models, the information matrix does not depend on the parameters $\boldsymbol{\theta}$, and is given by

$$\boldsymbol{M}(\xi) = \boldsymbol{X}' \boldsymbol{W} \boldsymbol{X}, \tag{2.5}$$

where $\boldsymbol{W} = \mathrm{diag}(w_1, \ldots, w_n)$.

For models which are nonlinear in their parameters, the information matrix now depends on $\boldsymbol{\theta}$, and is calculated by

$$\boldsymbol{M}(\boldsymbol{\theta}, \xi) = \boldsymbol{F}' \boldsymbol{W} \boldsymbol{F}, \tag{2.6}$$

where $\boldsymbol{F}$ is the $n \times p$ matrix of first partial derivatives:

$$\boldsymbol{F} = \left[\frac{\partial \eta(\boldsymbol{x}_1, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \ldots \quad \frac{\partial \eta(\boldsymbol{x}_n, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]'.$$

The information matrix for GLMs for binary data also depends on the model parameters, and is given by

$$\boldsymbol{M}(\boldsymbol{\theta}, \xi) = \boldsymbol{X}' \boldsymbol{W} \boldsymbol{X} \tag{2.7}$$

where $\boldsymbol{W}$ is now a function of more than just the design weights,

$$\boldsymbol{W} = \mathrm{diag}\left(\frac{w_i}{\pi_i(1-\pi_i)} \left(\frac{d\pi_i}{d\eta_i} \right)^2 \right).$$

The dependence of the information matrix on the parameter values is via the $\boldsymbol{W}$ matrix, whose elements are functions of π_i, which are in turn functions of $\boldsymbol{\theta}$. For example, the information matrix for a logistic model for binomial data involves the matrix $\boldsymbol{W} = \text{diag}(w_i\pi_i(1-\pi_i))$, where $\pi_i = \text{logit}^{-1}(\boldsymbol{x}_i'\boldsymbol{\theta})$.

2.3.2 Parameter estimation

D-optimality

The Cramer-Rao Lower Bound states that the variance-covariance matrix of any unbiased estimator of $\boldsymbol{\theta}$ is bounded below by $\boldsymbol{M}^{-1}(\boldsymbol{\theta}, \xi)$. It follows that any population design which maximises $|\boldsymbol{M}(\boldsymbol{\theta}, \xi)|$ (where $|\boldsymbol{A}|$ denotes the determinant of the matrix $\boldsymbol{A}$) minimises the volume of the joint asymptotic confidence ellipsoid for the parameters. Such a design is called D-optimal.

The ability of a design to efficiently estimate the parameters of a model can be addressed by its efficiency compared to the D-optimal design, i.e. its D-efficiency,

$$D_{\text{eff}}(\xi, \boldsymbol{\theta}) = \left(\frac{|\boldsymbol{M}(\boldsymbol{\theta}, \xi)|}{|\boldsymbol{M}(\boldsymbol{\theta}, \xi^{*D})|} \right)^{1/p}$$

where ξ^{*D} is the D-optimal design and p is the number of parameters in the model. More generally, we can compare two designs ξ_1 and ξ_2 using the relative efficiency of one compared to the other:

$$\text{Eff}(\xi_1, \xi_2, \boldsymbol{\theta}) = \left(\frac{|\boldsymbol{M}(\boldsymbol{\theta}, \xi_1)|}{|\boldsymbol{M}(\boldsymbol{\theta}, \xi_2)|} \right)^{1/p}$$

Other criteria

A large number of alphabetic optimality criteria exist for the creation of locally optimal designs for the estimation of parameters, such as: A, where the sum of the variances of the parameter estimates is minimised (by minimising the trace of the inverse of the information matrix); c, in which the variance of a linear combination of the parameters is minimised; G, which minimises the maximum over the design space of the standardised variance (the variance of the predicted response scaled for sample size and residual error variance); and D_s, where the interest is in estimating a subset s of the parameters as precisely as possible.

See Atkinson and Donev (1992), for example, for a review of these criteria and more. This thesis will be mainly concerned with the use of D-optimality for parameter estimation.

2.3.3 Model discrimination

T-optimality

Suppose that the response can be described by the nonlinear model in Equation (2.3), where the true function η is either of two candidate functions $\eta_1(\boldsymbol{x}, \boldsymbol{\theta}_1)$ and $\eta_2(\boldsymbol{x}, \boldsymbol{\theta}_2)$ with $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\theta}_r \in \Theta_r$. The construction of T-optimal designs, as described by Atkinson and Fedorov (1975a), depends on the assumption that one of these models is true, and that its parameters values are known. If the first model is arbitrarily chosen as true (and therefore $\boldsymbol{\theta}_1$ is assumed to be known), and we write

$$\eta_t(\boldsymbol{x}) = \eta_1(\boldsymbol{x}, \boldsymbol{\theta}_1),$$

then the T-optimal design can be defined, using notation similar to that of Uciński and Bogacka (2002), as

$$\xi^{*T} = \arg\max_{\xi \in \Xi} \left\{ \min_{\boldsymbol{\theta}_2 \in \Theta_2} J(\xi, \boldsymbol{\theta}_2) \right\}, \tag{2.8}$$

where

$$J(\xi,\, \boldsymbol{\theta}_2) = \sum_{i=1}^{n} w_i \|\eta_t(\boldsymbol{\xi}_i) - \eta_2(\boldsymbol{\xi}_i, \boldsymbol{\theta}_2)\|^2, \tag{2.9}$$

where $\|\cdot\|$ is the Euclidean norm and w_i are the design weights. Write

$$\boldsymbol{\theta}_2^{*T} = \arg\min_{\boldsymbol{\theta}_2 \in \Theta_2} J(\xi^{*T}, \boldsymbol{\theta}_2).$$

As noted by Atkinson and Fedorov (1975a), there are certain pairs of models (eg. nested models) for which it is meaningless to design these types of experiments when parameter values are unrestricted. In such cases it is necessary to place further constraints on $\boldsymbol{\theta}_2$.

Ponce de Leon and Atkinson (1992) defined the analogous T-optimality criterion for GLMs. Assume again that we have two candidate models for binary data, this time defined by their linear predictors $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, and their link functions g_1 and g_2, with the first model assumed to be true and its parameter values known. A T-optimal design in this case maximises the deviance arising from the fit of model 2 when the data are the expected

responses of model 1. That is, the T-optimality criterion is the deviance $D(\boldsymbol{\pi}_1; \hat{\boldsymbol{\pi}}_2)$, where $D(\cdot)$ is the function defined in Equation 2.4.

2.4 Search methods

The criteria outlined in Section 2.3 are used in a search of the design space in order to obtain the optimal experimental design. These criteria often form a surface over the design region with many local extrema. It is therefore important to use a global search method which performs a thorough search of the design space to avoid 'getting stuck' in local optima. Two such methods which have proved to be useful for optimal design in the context of this thesis are presented here. As the problems in this thesis are concerned with optimisation of continuous variables (such as sampling times and dose levels) rather than discrete variables, the continuous version of each algorithm is described here (both methods have evolved from combinatorial optimisation methods).

2.4.1 Simulated annealing

A search routine employing simulated annealing (SA) for continuous variables can be used to maximise the various design criteria. The implementation used in this thesis is based on the algorithm of Corana *et al.* (1987). Also see Goffe *et al.* (1994) for examples of its use. A brief overview of the procedure is given here.

Consider the design ξ in Equation (2.1) as an $(m+1) \times n$ matrix with elements $\xi_{i,j}$. In a constrained optimisation (the only type considered in this thesis), each element of this design matrix has a corresponding lower and upper bound. Write the matrices defining these bounds as $\boldsymbol{L}$ and $\boldsymbol{U}$. Since the $(m+1)$th row corresponds to the design weights, we always have $\boldsymbol{L}_{(m+1),j} = 0$ and $\boldsymbol{U}_{(m+1),j} = 1$ for $j = 1, \ldots, n$.

At the kth iteration of the algorithm, each element of ξ (i.e. each of the support points and the weights) is perturbed in sequence to give $\xi_{i,j}^k$. Thus each iteration involves finding $(m+1)n$ new designs. The size of the perturbations are defined by a matrix $\boldsymbol{V} = \{v_{i,j}\}$ of step sizes. The value of $\xi_{i,j}^k$ is taken from a uniform distribution on the interval $\xi_{i,j}^{k-1} \pm v_{i,j}$.

The $v_{i,j}$ are adjusted at specified intervals so that, on average, around half of the new designs are rejected according to the rule described below.

For each new design, its criterion $C(\xi^k)$ is calculated and compared to the criterion for the current 'best' design ξ^*. Designs with higher criteria are always accepted, designs with lower criteria are accepted or rejected according to the Metropolis criterion, i.e. the acceptance probability is $\exp[(C(\xi^k) - C(\xi^*))/T_k]$ for the current temperature T_k. The temperature is lowered at specified intervals by geometric cooling, $T_{k+1} = \alpha T_k$ ($\alpha = 0.9$ for these examples). The initial temperature is calculated after an initial run (of usually 100 or more iterations), with T_0 usually chosen such that the initial probability of accepting 'inferior' designs is close to 95%. Designs containing points outside the ranges defined by $\boldsymbol{L}$ and $\boldsymbol{U}$ are always rejected.

The algorithm stops when the average step size has reached a suitable value, i.e. the stopping criterion is $\sum_{i,\,j} v_{i,j}/[(m+1)n] < \text{tol}$, for a pre-specified small tolerance, tol. This differs slightly from the algorithm of Corana *et al.* (1987), who instead rely on the criterion reaching a stable value, regardless of the step sizes. Sufficiently small step sizes ensure that the criterion does in fact reach a stable value.

MATLAB code is given in Appendix A.1 for the algorithm as used in design optimisation problems in this thesis.

2.4.2 Cross-entropy

The cross-entropy (CE) method (Rubinstein and Kroese, 2004) was originally developed as an algorithm for estimating probabilities of rare events in stochastic networks. It was then adapted to, among other applications, combinatorial optimisation and, more recently, to continuous multi-extremal optimisation (Kroese *et al.*, 2005).

Again consider the design ξ as an $(m+1)\times n$ matrix with elements $\xi_{i,j}$, with corresponding lower and upper bounds given by the matrices $\boldsymbol{L}$ and $\boldsymbol{U}$. To apply the CE method to our optimal design problems more easily, we reshape each of these three matrices to $(m+1)n \times 1$ vectors. We associate the elements of ξ with independent truncated normal distributions, with lower and upper limits defined by the elements of $\boldsymbol{L}$ and $\boldsymbol{U}$. The means and variances

of these distributions are given by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{(m+1)n})'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_{(m+1)n}^2)'$. The truncated normal distribution for each element of ξ is then written as $\xi_j \sim \mathrm{N}_t(\mu_j, \sigma_j^2, \boldsymbol{L}_j, \boldsymbol{U}_j)$, for $j = 1, \ldots, (m+1)n$. We are interested in finding the point $\boldsymbol{\mu}^*$ which relates to the optimal design ξ^*.

We begin by choosing initial estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, written $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\sigma}}_0$. Usually $\hat{\boldsymbol{\mu}}_0$ is chosen as the centre of the rectangular design region, $\boldsymbol{L} + (\boldsymbol{U} - \boldsymbol{L})/2$, and $\hat{\boldsymbol{\sigma}}_0$ is $(\boldsymbol{U} - \boldsymbol{L})/4$. For the kth iteration of the algorithm, we draw a random sample $\xi^1, \ldots, \xi^N$ from the $\mathrm{N}_t(\hat{\boldsymbol{\mu}}_{k-1}, \hat{\boldsymbol{\sigma}}_{k-1}^2, \boldsymbol{L}, \boldsymbol{U})$ distribution. Let $\mathcal{I}$ be the indices of the N^{elite} ($= \rho N$, typically $\rho = 0.1$) best performing samples, as chosen using the optimality criterion. For $j = 1, \ldots, (m+1)n$, we let

$$\tilde{\mu}_{kj} = \sum_{i \in \mathcal{I}} \xi_j^i / N^{\text{elite}}$$

and

$$\tilde{\sigma}_{kj}^2 = \sum_{i \in \mathcal{I}} (\xi_j^i - \tilde{\mu}_{kj})^2 / N^{\text{elite}}.$$

Finally, the parameters are smoothed by the following:

$$\hat{\boldsymbol{\mu}}_k = \alpha_1 \tilde{\boldsymbol{\mu}}_k + (1 - \alpha_1)\hat{\boldsymbol{\mu}}_{k-1}, \quad \hat{\boldsymbol{\sigma}}_k^2 = \alpha_2 \tilde{\boldsymbol{\sigma}}_k^2 + (1 - \alpha_2)\hat{\boldsymbol{\sigma}}_{k-1}^2,$$

where α_1 and α_2 are the fixed smoothing parameters, with typically $\alpha_1 = 0.9$ and $\alpha_2 = 0.3$. The algorithm converges when $\max_j(\hat{\sigma}_{kj}) < \text{tol}$ for some fixed tol (tol $= 0.01$ is used in the problems in this thesis).

Additionally, we use the 'injection' method described in Botev and Kroese (2004), in which every time the stopping criterion is met, the standard deviations are inflated by adding

$$|C_k^* - C_{k-1}^*| h$$

for some h between 0.1 and 10, where C_k^* is the best value of the criterion obtained at iteration k. The whole process is then repeated a number of times. This is used to avoid settling into inferior local optima.

MATLAB code is given in Appendix A.2 for the algorithm as used in design optimisation problems in this thesis.

2.4.3 Comparison of algorithms

A direct comparison of the two algorithms described above, in terms of their run times and ability to find the global optimum, is difficult, considering the need for quite a large number of parameters to be controlled for each algorithm. In each case we could set these parameters so that the algorithm runs for a long time, to be confident of finding a global optimum. For example, in the simulated annealing algorithm we could set the number of iterations between step size changes to be very large, and the stopping criterion 'tol' to be very small. This may virtually guarantee that the global optimum is found, but it is possible that a much shorter search will produce the same design.

As a typical example of optimisation problems tackled in this thesis, a comparison of the two methods was carried out by finding D-optimal designs using the logistic regression model described in Section 6.4.2. The default values of the algorithm parameters (as defined in the code in Appendix A) were used in both cases, except that no 'injections' were used in the cross-entropy method. Both algorithms converged to the same design (as given in Section 6.4.2), but SA was slightly slower in this case, converging in 144 seconds, as opposed to 138 seconds for CE.

It is encouraging that the two methods converged to the same design, but we cannot make strong judgements about the computation time. The slower algorithm in this case, SA, was run again with the numbers of cycles between changes in step size and temperature reduced from the default values of 20 and 10 to 10 and 5, respectively. The algorithm converged to the same solution, but this time in 79 seconds, considerably faster than the first run, even considerably faster than the first run of the cross-entropy method. However, we may be able to adjust the parameters of the CE algorithm to obtain the same result in even faster time, and so on.

In most of the problems in this thesis, rather than aiming for the algorithm which has the greatest speed, we are instead more concerned with finding the global optimum. Simulated annealing seems to consistently find the global optimum designs for these default values of the algorithm's parameters, regardless of the models under consideration. Reasonable values of the parameters for the cross-entropy method, on the other hand, seem to be much more

parameter dependent. In light of this, simulated annealing is used for most problems in this thesis, with CE used as a 'backup' to check for global optimality when practical.

Part I

Nonlinear models

This part of the thesis investigates methods of optimal design for nonlinear models when the objective is to both effectively discriminate between two competing models and efficiently estimate the parameters of both models. A number of methods involving multiple criteria are introduced, and their use is demonstrated in a number of examples. An example of a real-life optimal design is also given where the objective is both discrimination and estimation. The models are extremely complex, involving random effects, multiple responses, and systems of ordinary differential equations with no analytic solution.

Chapter 3

Designs for discrimination and estimation in nonlinear models

Research into the optimal design of experiments has in the main concentrated on optimisation with respect to parameter estimation. An experimental design is 'optimised', that is the best choice of runs, experimental units, etc. is found through the use of an optimality criterion. The most commonly used criterion is D-optimality, described in Section 2.3.2, which minimises the determinant of the variance-covariance matrix of the parameter estimates. In comparison, relatively little research has been carried out into criteria and techniques for discriminating between competing models for an experimental design. Atkinson and Fedorov (1975a,b) introduced T-optimality as a criterion for discriminating between models. T-optimality has not been widely used, due in part to the computational burden involved in the optimisation for even relatively simple models, but also due to the theoretical challenges it presents (see Uciński and Bogacka (2002)).

It is unclear as to whether T-optimal designs are good D-optimal designs and vice versa. It would be very useful and advantageous to be able to have designs that perform well for both model discrimination and model parameter estimation simultaneously. Compound criteria, such as the product of D-optimality criteria (Atkinson and Cox, 1974) have been suggested, and this is followed up on here. Other uses of composite or compromise criteria for optimal design are described in Stigler (1971), Studden (1982) and Cook and Wong (1994).

In this chapter, criteria and methods are introduced which are aimed at achieving designs that efficiently discriminate between models and yield efficient parameters estimates. Product, conditional and hybrid optimality criteria are developed and applied to both linear and nonlinear models. The resulting designs are evaluated in terms of both model discrimination and parameter estimation.

3.1 Optimality criteria for two competing models

Compound or composite criteria, such as those suggested by Atkinson and Cox (1974), are usually aimed at combining two (say) design properties in which we are interested. As mentioned above it is not known whether a design can be simultaneously T- and D-optimal. The product optimality criterion based on the work of Atkinson and Cox is defined in this section, along with several alternative criteria which are aimed at achieving designs with some degree of effective model discrimination and efficient parameter estimation. Each of the proposed criteria use either or both T- or D-optimality objective functions as defined in Section 2.3.

In this chapter we only consider approximate designs, i.e. designs of the form

$$\xi = \begin{Bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix},$$

as described in Section 2.1. Optimal designs presented here are written

$$\xi^* = \begin{Bmatrix} \boldsymbol{x}^* \\ \boldsymbol{w}^* \end{Bmatrix},$$

where $\boldsymbol{x}^* = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ and $\boldsymbol{w}^* = (w_1, \dots, w_n)$.

Suppose that the rth of our two competing models for the observations can be given by

$$y_i = \eta_r(\boldsymbol{\xi}_i, \boldsymbol{\theta}_r) + \varepsilon_i \quad (i = 1, \dots, n;\ r = 1, 2),$$

with $\varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, and write $\boldsymbol{Y}$ as the $n \times 1$ vector of responses and $\boldsymbol{\theta}_r = (\theta_{r1}, \dots, \theta_{rn_r})'$. As we are considering T-optimal designs, among others, in this chapter, we require that one

of the two models is assumed to be true. Suppose that the first model is true (and therefore assume that $\boldsymbol{\theta}_1$ is known), and write

$$\eta_t(\boldsymbol{\xi}_i) = \eta_1(\boldsymbol{\xi}_i, \boldsymbol{\theta}_1).$$

3.1.1 T-optimality and T-efficiency

Recall from Section 2.3.3 that the T-optimal design is defined as

$$\xi^{*T} = \arg\max_{\xi \in \Xi} \left\{ \min_{\boldsymbol{\theta}_2 \in \Theta_2} J(\xi, \boldsymbol{\theta}_2) \right\}, \tag{3.1}$$

where

$$J(\xi, \boldsymbol{\theta}_2) = \sum_{i=1}^{n} w_i \|\eta_t(\boldsymbol{\xi}_i) - \eta_2(\boldsymbol{\xi}_i, \boldsymbol{\theta}_2)\|^2, \tag{3.2}$$

where $\|\cdot\|$ is the Euclidean norm and w_i are the design weights. We have also defined $\boldsymbol{\theta}_2^{*T} = \arg\min_{\boldsymbol{\theta}_2 \in \Theta_2} J(\xi^{*T}, \boldsymbol{\theta}_2)$.

A comparison of sequentially constructed designs to the T-optimal design is given for a pair of linear models in Atkinson and Donev (1992, p.235). The 'efficiency' of a design ξ in terms of model discrimination used here is the ratio

$$\frac{J(\xi, \tilde{\boldsymbol{\theta}}_2)}{J(\xi^{*T}, \boldsymbol{\theta}_2^{*T})}.$$

However, the choice of $\tilde{\boldsymbol{\theta}}_2$ in the numerator presents a problem which is not addressed in Atkinson and Donev (1992). The T-optimal design gives the greatest separation between the response of the two models for the 'worst-case scenario' of $\boldsymbol{\theta}_2$, so it makes sense to assess the new design ξ in a similar manner, i.e. let $\tilde{\boldsymbol{\theta}}_2 = \hat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}_2 \in \Theta_2} J(\xi, \boldsymbol{\theta}_2)$. On the other hand, the competing D-optimal designs for nonlinear models rely on the prior specification of $\tilde{\boldsymbol{\theta}}_2 = \boldsymbol{\theta}_2^{*T}$, the same value used in the T-optimal design. One may argue that in this case it is fair to judge the design using the same assumption of parameter values as was used to find the design.

In this light, both methods of T-efficiencies, termed T_{eff}^a and T_{eff}^b, are included in this chapter:

$$T_{\text{eff}}^a(\xi) = \frac{J(\xi, \hat{\boldsymbol{\theta}}_2)}{J(\xi^{*T}, \boldsymbol{\theta}_2^{*T})}, \quad T_{\text{eff}}^b(\xi) = \frac{J(\xi, \boldsymbol{\theta}_2^{*T})}{J(\xi^{*T}, \boldsymbol{\theta}_2^{*T})}, \tag{3.3}$$

where $\hat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}_2 \in \Theta_2} J(\xi, \boldsymbol{\theta}_2)$.

3.1.2 Product optimality

If the objective of our experiment is not model discrimination, but efficient estimation of both sets of parameters of two competing models, then we may wish to maximise the product of the two D-optimality criteria, scaled for the number of parameters, as suggested by Atkinson and Cox (1974). The concept and definition of D-optimal designs were introduced in Section 2.3.2. The notation is now extended to account for the inclusion of multiple models in the design process.

The product optimal design for the two models is defined as

$$\xi^{*P} = \arg\max_{\xi\in\Xi} |\boldsymbol{M}_1(\boldsymbol{\theta}_1, \xi)|^{1/p_1} |\boldsymbol{M}_2(\boldsymbol{\theta}_2, \xi)|^{1/p_2}, \tag{3.4}$$

where $\boldsymbol{M}_r(\boldsymbol{\theta}_r, \xi)$ is the information matrix for the rth model and $\boldsymbol{\theta}_r$ is the vector of the p_r parameters in the model. Here equal importance is placed on each model. The exponents $1/p_1$ and $1/p_2$ may be adjusted if we wish to estimate the parameters of one model more efficiently than the other. For example, if we were twice as interested in estimating the parameters in model 1 as model 2 (or if we had twice as much confidence in model 1 as model 2), we might use the exponents $2/p_1$ and $1/p_2$ in Equation 3.4 instead.

If the models in question are nonlinear in their parameters, then the information matrix depends on the parameter vectors, as described in Section 2.3.1. Hence, to find product optimal designs the values of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ must first be specified. The designs described in the following sections combine this product criterion with the T-optimality criterion. For T-optimal designs, it is already assumed that the value of $\boldsymbol{\theta}_1$ is known, so the same $\boldsymbol{\theta}_1$ is used for the construction of these product optimal designs. The value of $\boldsymbol{\theta}_2$, however, is unknown, and we let it take the value obtained in finding the T-optimal design, i.e. we let $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^{*T}$.

The ability of a design to estimate the parameters of the rth model can be addressed by its efficiency compared to the D-optimal design, i.e. its D-efficiency,

$$D^r_{\text{eff}}(\boldsymbol{\theta}_r, \xi) = \left(\frac{|\boldsymbol{M}_r(\boldsymbol{\theta}_r, \xi)|}{|\boldsymbol{M}_r(\boldsymbol{\theta}_r, \xi^{*Dr})|} \right)^{1/p_r}, \quad r = 1, 2,$$

where ξ^{*Dr} is the D-optimal design, which maximises $|\boldsymbol{M}_r(\boldsymbol{\theta}_r, \xi)|$.

3.1.3 Conditional optimality

Since the construction of T-optimal designs does not take parameter estimation into account, it could be expected that the D-efficiencies of the designs may be quite poor in practice, a problem which is investigated in the examples which follow. In order to find designs which are useful for both parameter estimation *and* model discrimination, methods are proposed involving either the extension of the T-optimal design to include additional support points so that the overall design maximises the product criterion in Equation (3.4); or conversely the extension of the product optimal design to include points so that the overall design maximises the T-optimality criterion.

Suppose that we have derived a T-optimal design,

$$\xi^{*T} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*T} & \boldsymbol{\xi}_2^{*T} & \dots & \boldsymbol{\xi}_{n_T}^{*T} \\ w_1^{*T} & w_2^{*T} & \dots & w_{n_T}^{*T} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{x}^{*T} \\ \boldsymbol{w}^{*T} \end{Bmatrix}.$$

We wish to find m further support points which aid in parameter estimation to add to the existing design. Of course, the weights $w_1^{*T}, \dots, w_{n_T}^{*T}$ will need to be adjusted in order to 'make room' for the additional points (all weights must sum to 1). To do this, simply multiply each weight by a factor $(1-\alpha)$, where $\alpha \in (0, 1)$ represents the importance the experimenter places on parameter estimation. The construction of the 'conditional' optimal design (conditional on the fixed T-optimal design) then involves finding the additional support points $\boldsymbol{\xi}_{n_T+1}^{*P|T}, \dots, \boldsymbol{\xi}_{n_T+m}^{*P|T}$ and weights $w_{n_T+1}^{*P|T}, \dots, w_{n_T+m}^{*P|T}$ so that the complete design

$$\xi_\alpha^{*P|T} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*T} & \dots & \boldsymbol{\xi}_{n_T}^{*T} & \boldsymbol{\xi}_{n_T+1}^{*P|T} & \dots & \boldsymbol{\xi}_{n_T+m}^{*P|T} \\ (1-\alpha)w_1^{*T} & \dots & (1-\alpha)w_{n_T}^{*T} & w_{n_T+1}^{*P|T} & \dots & w_{n_T+m}^{*P|T} \end{Bmatrix}$$

maximises the product criterion from Equation (3.4), and

$$\sum_{i=1}^{n_T}(1-\alpha)w_i^{*T} + \sum_{i=n_T+1}^{n_T+m} w_i^{*P|T} = 1.$$

Denote this design by $\xi_\alpha^{*P|T}$ to indicate that the T-optimal design is found first, followed by additional 'product optimal' points (with a weighting of α on product optimality), and the design shall be referred to as $P|T$-optimal.

As already noted, for non-linear models these information matrices will depend on the parameter values $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. In other words, the conditional optimal design will be locally

optimal. As in the product optimal design, let $\boldsymbol{\theta}_1$ take the value already assumed to be known, and for the second model, let the parameters be dictated by the T-optimal design: let $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^{*T}$.

The obvious reversal of this process is fix the points of the product optimal design

$$\xi^{*P} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*P} & \boldsymbol{\xi}_2^{*P} & \dots & \boldsymbol{\xi}_{n_P}^{*P} \\ w_1^{*P} & w_2^{*P} & \dots & w_{n_P}^{*P} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{x}^{*P} \\ \boldsymbol{w}^{*P} \end{Bmatrix},$$

with weights multiplied by a factor α (again representing the importance placed on parameter estimation), and to choose m additional support points such that the complete design

$$\xi_\alpha^{*T|P} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*P} & \dots & \boldsymbol{\xi}_{n_P}^{*P} & \boldsymbol{\xi}_{n_P+1}^{*T|P} & \dots & \boldsymbol{\xi}_{n_P+m}^{*T|P} \\ \alpha w_1^{*P} & \dots & \alpha w_{n_P}^{*P} & w_{n_P+1}^{*T|P} & \dots & w_{n_P+m}^{*T|P} \end{Bmatrix}$$

maximises the T-optimality criterion from Equation (2.8), where

$$\sum_{i=1}^{n_P} \alpha w_i^{*P} + \sum_{i=n_P+1}^{n_P+m} w_i^{*T|P} = 1.$$

Denote this design by $\xi_\alpha^{*T|P}$ to indicate that the product optimal design is found first, followed by additional 'T-optimal' points (with a weighting of α on product optimality), and we shall refer to the design as $T|P$-optimal.

3.1.4 Hybrid designs

A more straightforward approach to constructing designs which address both problems of discrimination and estimation is to simply include design points from both the T-optimal design and the product optimal design in such a way that the weight from each point is adjusted to account for the priority given to the criterion which that point represents. For the n_P-point product optimal design ξ^{*P} and the n_T-point T-optimal design ξ^{*T}, and for an $\alpha \in (0, 1)$ again representing the importance placed on parameter estimation, such a design is given by

$$\xi_\alpha^{\text{hybrid}} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*P} & \dots & \boldsymbol{\xi}_{n_P}^{*P} & \boldsymbol{\xi}_1^{*T} & \dots & \boldsymbol{\xi}_{n_T}^{*T} \\ \alpha w_1^{*P} & \dots & \alpha w_{n_P}^{*P} & (1-\alpha) w_1^{*T} & \dots & (1-\alpha) w_{n_T}^{*T} \end{Bmatrix}.$$

Note that, by definition,

$$\sum_{i=1}^{n_P} \alpha w_i^{*P} + \sum_{i=1}^{n_T} (1-\alpha) w_i^{*T} = 1.$$

This approach has the advantage that it does not rely on further optimisation once the T-optimal and product optimal designs are known.

3.2 Construction of designs and examples

3.2.1 Algorithms

The examples that follow involve the numerical optimisation of the criteria presented in Section 3.1. The algorithm used in each case is an implementation of simulated annealing for continuous variables, as described in Section 2.4.1.

In terms of the T-optimal designs, a special note must be made of the minimisation of the function in Equation (3.2): at each iteration of the annealing algorithm, this was implemented using MATLAB's Optimisation Toolbox. The functions `fminsearch` and `lsqnonlin` may both be used for this purpose. `fminsearch` minimises any scalar function of several variables, whereas `lsqnonlin` is designed for minimisations of the form

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \sum_{i=1}^{q} f_i(\boldsymbol{x})^2 + L$$

where L is a constant. Depending on the model structure and the initial estimate of $\boldsymbol{\theta}_2$, for some designs `fminsearch` found the best $\boldsymbol{\theta}_2$ out of the two functions (i.e. it gave a better fit for for model 2), but for other designs `lsqnonlin` was better. Thus the criterion calculated may be larger than the true value of the criterion, which results in the algorithm converging to less than optimal designs. Due to this inconsistency it was necessary to use both functions to perform the minimisation, after which the smallest minimum was selected. This heuristic approach appears to find a stable minimum and therefore avoids the problem with convergence.

3.2.2 Linear example

In this first example, the results of Example 20.3 of Atkinson and Donev (1992) are reproduced, and then product, conditional and hybrid optimal designs are found. The two models which we wish to discriminate between are

$$\eta_1(x, \boldsymbol{\theta}_1) = \theta_{11} + \theta_{12}e^x + \theta_{13}e^{-x} \quad (-1 \leq x \leq 1, -\infty \leq \theta_{1j} \leq \infty),$$

and

$$\eta_2(x, \boldsymbol{\theta}_2) = \theta_{21} + \theta_{22}x + \theta_{23}x^2 \quad (-1 \leq x \leq 1, -\infty \leq \theta_{2j} \leq \infty).$$

The first model is assumed to be true for the construction of the T-optimal design, with parameter values $\theta_{11} = 4.5$, $\theta_{12} = -1.5$ and $\theta_{13} = -2$, as in Atkinson and Donev (1992). The T-optimal design (identical to the one found by Atkinson and Donev), product optimal design, conditional designs (with either the T-optimal or product optimal design first, for a range of α), hybrid designs (also for a range of α), and their D- and T-efficiencies are given in Table 3.1.

The support points of the $T|P$ optimal designs seem to be identical to those of the hybrid design, whereas the $P|T$-optimal design is slightly different. It should be noted that in terms of T^b_{eff}, the conditional designs and hybrid designs are all very similar. The $P|T$-optimal designs are worse than the $T|P$-optimal and hybrid designs in terms of T^b_{eff}, but more D-efficient.

The efficiencies for a range of conditional and hybrid optimal designs, including the product design ($\alpha = 1$) and the T-optimal design ($\alpha = 0$) are shown graphically in Figure 3.1. The three methods all produce very similar results. There is an obvious increasing trend in D-efficiencies of all designs as we place more importance on parameter estimation (i.e. as α decreases) and an almost linear increase in both T-efficiencies as we favour model discrimination (i.e. as α increases). A reasonable trade-off between the two objectives seems to be at around $\alpha = 0.2$ for the $P|T$-optimal design, and at around $\alpha = 0.35$ for the other optimal designs.

Design, $\xi^* = \begin{Bmatrix} \boldsymbol{x}^* \\ \boldsymbol{w}^* \end{Bmatrix}$	Efficiency			
	D^1_{eff}	D^2_{eff}	T^a_{eff}	T^b_{eff}
T-optimal $\xi^{*T} = \begin{Bmatrix} -1 & -0.6693 & 0.1439 & 0.9569 \\ 0.2527 & 0.4281 & 0.2469 & 0.0723 \end{Bmatrix}$	0.6069	0.6179	1	1
D-optimal, model 1 $\xi^{*D_1} = \begin{Bmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$	1	1	0	0.9082
D-optimal, model 2 $\xi^{*D_2} = \begin{Bmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$	1	1	0	0.9082
D-optimal, product $\xi^{*P} = \begin{Bmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$	1	1	0	0.9082
$P\|T$-optimal $\xi^{*P\|T}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*T} & -1 & 0.0682 & 1 \\ 0.75\boldsymbol{w}^{*T} & 0.0126 & 0.0178 & 0.2196 \end{Bmatrix}$	0.8340	0.8407	0.8193	0.9885
$\xi^{*P\|T}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*T} & -1 & 0.0448 & 1 \\ 0.5\boldsymbol{w}^{*T} & 0.1172 & 0.1209 & 0.2619 \end{Bmatrix}$	0.8963	0.9012	0.6928	0.9740
$\xi^{*P\|T}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*T} & -1 & 0.0218 & 1 \\ 0.25\boldsymbol{w}^{*T} & 0.2244 & 0.2261 & 0.2996 \end{Bmatrix}$	0.9511	0.9537	0.4234	0.9479
$T\|P$-optimal $\xi^{*T\|P}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*P} & -1 & -0.6693 & 0.1438 & 0.9570 \\ 0.25\boldsymbol{w}^{*P} & 0.1896 & 0.3211 & 0.1852 & 0.0542 \end{Bmatrix}$	0.7648	0.7721	0.9058	0.9770
$\xi^{*T\|P}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*P} & -1 & -0.6693 & 0.1438 & 0.9570 \\ 0.5\boldsymbol{w}^{*P} & 0.1264 & 0.2141 & 0.1234 & 0.0361 \end{Bmatrix}$	0.8691	0.8740	0.7134	0.9541
$\xi^{*T\|P}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*P} & -1 & -0.6693 & 0.1437 & 0.9570 \\ 0.75\boldsymbol{w}^{*P} & 0.0632 & 0.1070 & 0.0617 & 0.0181 \end{Bmatrix}$	0.9448	0.9473	0.4213	0.9311
Hybrid $\xi^{\text{hybrid}}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.25\boldsymbol{w}^{*P} & 0.75\boldsymbol{w}^{*T} \end{Bmatrix}$	0.7648	0.7721	0.9058	0.9770
$\xi^{\text{hybrid}}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.5\boldsymbol{w}^{*P} & 0.5\boldsymbol{w}^{*T} \end{Bmatrix}$	0.8691	0.8740	0.7134	0.9541
$\xi^{\text{hybrid}}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.75\boldsymbol{w}^{*P} & 0.25\boldsymbol{w}^{*T} \end{Bmatrix}$	0.9448	0.9473	0.4213	0.9311

Table 3.1: Efficiencies of near-optimal designs for two competing linear models.

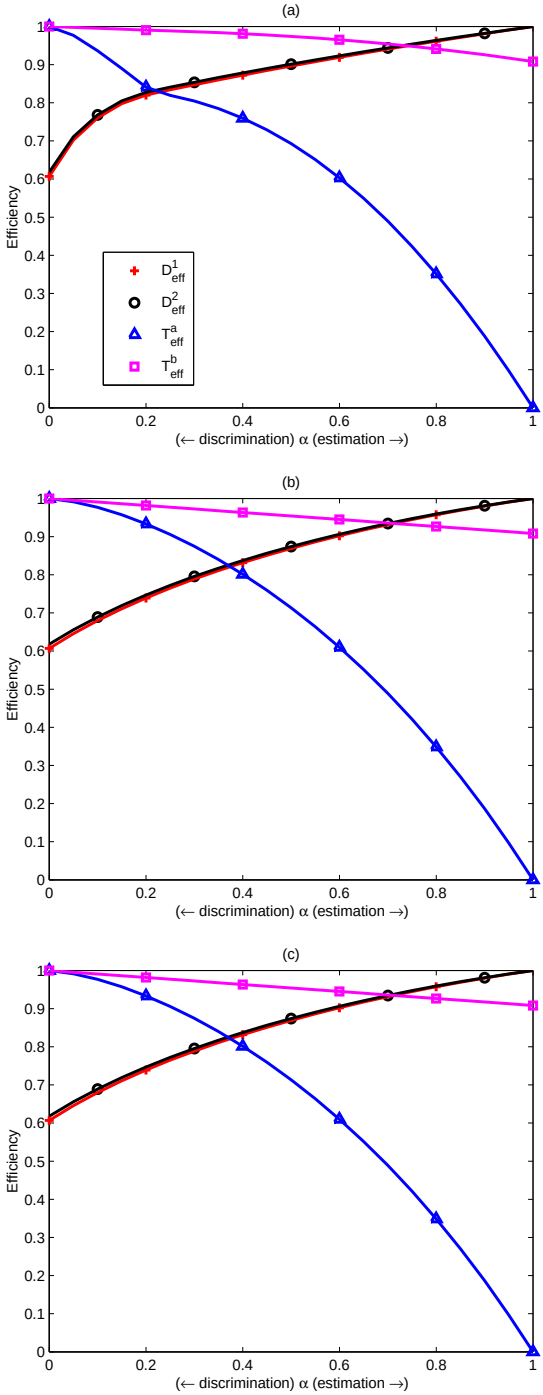

FIGURE 3.1: Efficiencies of designs for two competing linear models: (a) $P|T$-optimal designs; (b) $T|P$-optimal designs; (c) hybrid designs. In each case, α represents the importance placed on parameter estimation.

3.2.3 Nonlinear example: two models for decay

Now consider the problem of finding an optimal design to discriminate between two models for decay as described in Example 20.1 of Atkinson and Donev (1992), namely the exponential decay model

$$\eta_1(x,\, \theta_1) = \exp(-\theta_1 x) \quad (x,\, \theta_1 \geq 0),$$

and the inverse polynomial

$$\eta_2(x,\, \theta_2) = \frac{1}{1+\theta_2 x} \quad (x,\, \theta_2 \geq 0).$$

Again let the first model be true, with $\theta_1 = 1$, as in Atkinson and Donev (1992). The same process as the previous example was followed to find the T-optimal design, the product optimal design and the conditional and hybrid optimal designs. The results are summarised in the same manner in Table 3.2 and Figure 3.2. (The T-optimal design is again confirmed by the results given for this example in Atkinson and Donev (1992).)

As for the designs for the linear models, the $T|P$-optimal and hybrid designs appear to be identical, whereas the $P|T$-optimal designs differ slightly.

Trends similar to those seen in the previous example are shown here, with values of α around 0.5 giving similar efficiencies for parameter estimation and model discrimination in each case. In this case, however, there seems to be a bigger trade-off between the T-optimal and product optimal designs than the trade-off in the linear models. That is, as α increases the D-efficiencies increase and the T-efficiencies decrease at a greater rate than for the linear models.

3.2.4 Nonlinear example: two models for rise and decay

We now consider a pair of slightly more complicated models, used in the case study of Section 3.13 of Bates and Watts (1988). The first is the three-parameter quadratic Michaelis-Menten type model

$$\eta_1(x,\, \boldsymbol{\theta}_1) = \frac{\theta_{11} x}{\theta_{12} + x + \theta_{13} x^2} \quad (x \geq 0,\, -\infty \leq \theta_{1j} \leq \infty),$$

and the second is the exponential difference model

$$\eta_2(x,\, \boldsymbol{\theta}_2) = \theta_{21}\left(e^{-\theta_{23} x} - e^{-\theta_{22} x}\right) \quad (x \geq 0,\, -\infty \leq \theta_{2j} \leq \infty).$$

Design, $\xi^* = \begin{Bmatrix} \boldsymbol{x}^* \\ \boldsymbol{w}^* \end{Bmatrix}$	Efficiency			
	D^1_{eff}	D^2_{eff}	T^a_{eff}	T^b_{eff}
T-optimal $\xi^{*T} = \begin{Bmatrix} 0.327 & 3.34 \\ 0.3345 & 0.6655 \end{Bmatrix}$	0.2065	0.4472	1	1
D-optimal, model 1 $\xi^{*D_1} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$	1	0.8224	0	0.0403
D-optimal, model 2 $\xi^{*D_2} = \begin{Bmatrix} 0.5324 \\ 1 \end{Bmatrix}$	0.7221	1	0	0.7326
D-optimal, product $\xi^{*P} = \begin{Bmatrix} 0.7995 \\ 1 \end{Bmatrix}$	0.9545	0.9212	0	0.2393
$P\|T$-optimal $\xi^{*P\|T}_{\alpha=0.25} = \begin{Bmatrix} x^{*T} & 0.850 \\ 0.75w^{*T} & 0.25 \end{Bmatrix}$	0.3986	0.5598	0.7787	0.7932
$\xi^{*P\|T}_{\alpha=0.5} = \begin{Bmatrix} x^{*T} & 0.825 \\ 0.5w^{*T} & 0.5 \end{Bmatrix}$	0.5861	0.6783	0.5397	0.6022
$\xi^{*P\|T}_{\alpha=0.75} = \begin{Bmatrix} x^{*T} & 0.810 \\ 0.25w^{*T} & 0.75 \end{Bmatrix}$	0.7710	0.7992	0.2791	0.4187
$T\|P$-optimal $\xi^{*T\|P}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*P} & 0.327 & 3.34 \\ 0.25\boldsymbol{w}^{*P} & 0.2509 & 0.4991 \end{Bmatrix}$	0.3935	0.5657	0.7893	0.8098
$\xi^{*T\|P}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*P} & 0.327 & 3.34 \\ 0.5\boldsymbol{w}^{*P} & 0.1672 & 0.3328 \end{Bmatrix}$	0.5805	0.6842	0.5461	0.6196
$\xi^{*T\|P}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*P} & 0.327 & 3.34 \\ 0.75\boldsymbol{w}^{*P} & 0.0836 & 0.1664 \end{Bmatrix}$	0.7675	0.8027	0.2809	0.4295
Hybrid $\xi^{\text{hybrid}}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.25\boldsymbol{w}^{*P} & 0.75\boldsymbol{w}^{*T} \end{Bmatrix}$	0.3935	0.5657	0.7893	0.8098
$\xi^{\text{hybrid}}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.5\boldsymbol{w}^{*P} & 0.5\boldsymbol{w}^{*P} \end{Bmatrix}$	0.5805	0.6842	0.5461	0.6196
$\xi^{\text{hybrid}}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.75\boldsymbol{w}^{*P} & 0.25\boldsymbol{w}^{*P} \end{Bmatrix}$	0.7675	0.8027	0.2809	0.4295

Table 3.2: Efficiencies of near-optimal designs for two competing models for decay.

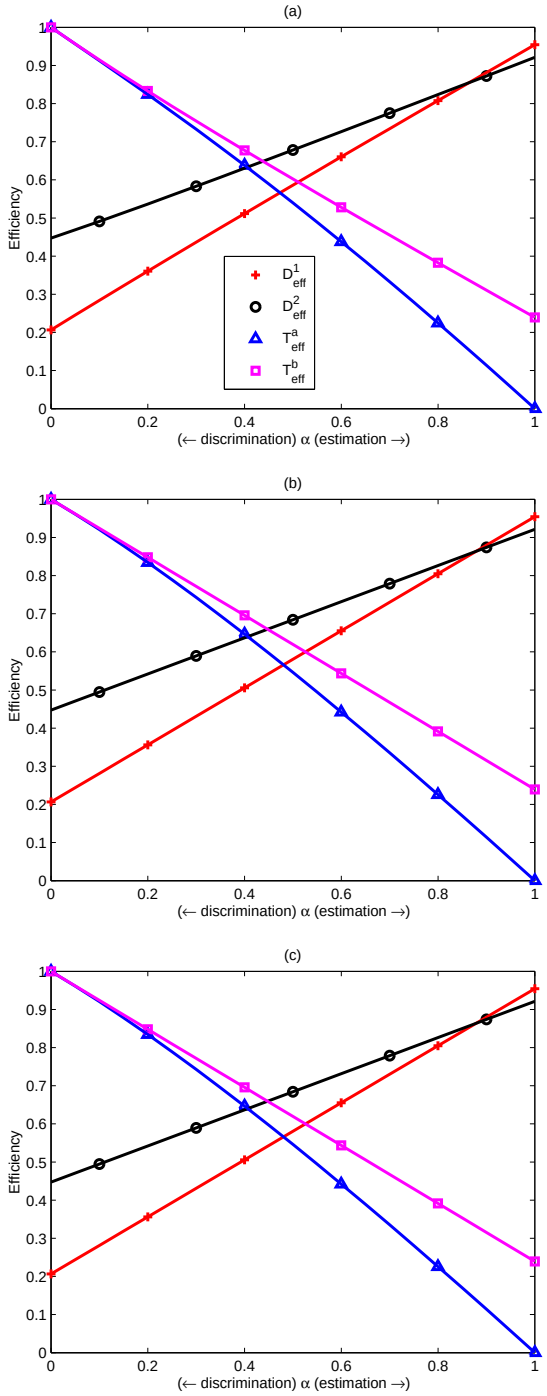

Figure 3.2: Efficiencies of designs for two models for decay: (a) $P|T$-optimal designs; (b) $T|P$-optimal designs; (c) hybrid designs. In each case, α represents the importance placed on parameter estimation.

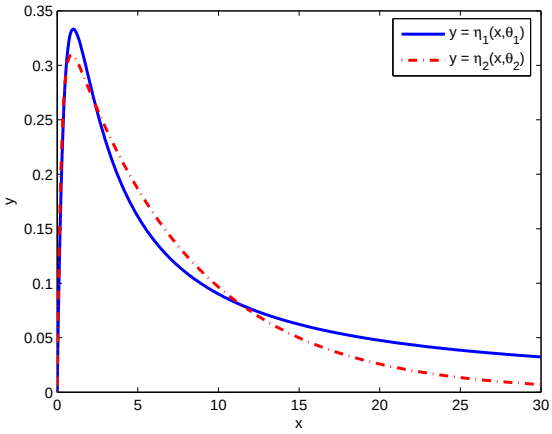

FIGURE 3.3: Two candidate models for rise and decay.

These can be seen as an extension of the models of the previous example, as they involve an initial rise in response, followed by a decay. Figure 3.3 shows the similarity of the models, despite their structural disparity.

Again, to find the T-optimal design, assume that the first model is true, and arbitrarily select $\theta_{11} = \theta_{12} = \theta_{13} = 1$. The resulting T-optimal design is given in Table 3.3, along with the product design and a selection of conditional and hybrid optimal designs.

Comparing the conditional and hybrid designs, the $T|P$-optimal designs appear yet again to be identical to the hybrid designs. The $P|T$-optimal designs again differ very slightly to the hybrid designs. It is interesting to note that in the $\alpha = 0.25$ case, where little emphasis is placed on parameter estimation, the $P|T$-optimal design contains only two additional points, whereas for higher values of α, three additional support points are needed to maximise the product criterion.

It can be seen again that the T-optimal design has a poor D-efficiency for each model and the D-optimal product design performs quite well in terms of parameter estimation and model discrimination when comparing T^b_{eff} (although not in terms of T^a_{eff}).

Figure 3.4 shows the decreasing trend in D-efficiencies and the increase in T-efficiencies as more importance is placed on model discrimination (i.e. as α increases). Again, this trade-off between designs is more pronounced for these nonlinear models than it is for the

linear models.

3.3 Discussion

We have seen that although T-optimal designs can be inefficient for parameter estimation, and conversely D-optimal designs may be unsuitable for model discrimination, we are able to construct designs using compound criteria which offer a suitable compromise between the two objectives. Through the conditional and hybrid designs described here, the trade-off between estimation and discrimination can be simply specified by the choice of α.

T^a_{eff} and T^b_{eff} always agree with respect to the direction of change. T^a_{eff}, however, always seems to show the greatest degree of change for a given change in α. Regardless of the choice of T-efficiency, the results are the same in terms of model discrimination: the lower the value of α (i.e. the more emphasis that is placed on model discrimination), the larger the T-efficiency of the conditional and hybrid designs.

The difference between the two types of conditional designs and the hybrid designs is minor. Indeed, for all three examples presented here, the conditional designs with the product optimal design points fixed are essentially equal to the corresponding hybrid designs. The use of hybrid designs is hence preferable, as they involve no further optimisation once the T-optimal and product optimal designs are known, and can be generated very quickly.

These results show that D-optimal designs (including product designs) with few support points offer little value in terms of discrimination. In these cases the additional support points in the conditional and hybrid designs give a dramatic improvement in T-efficiencies. It would appear that for models whose D-optimal designs have several support points (where the number of parameters in each model is larger), a product optimal design is efficient for both model discrimination and parameter estimation. Certainly more than a single support point would be required for effective model discrimination, regardless of the number of parameters in the models. Of course, a loss in parsimony (due to an increase in the number of distinct support points) may incur an additional cost in many practical situations. We may prefer to penalise the criteria presented here for less parsimonious designs.

The methods described in this chapter have been exemplified by the use of several sets

Design, $\xi^* = \left\{ \begin{matrix} \boldsymbol{x}^* \\ \boldsymbol{w}^* \end{matrix} \right\}$	Efficiency			
	D^1_{eff}	D^2_{eff}	T^a_{eff}	T^b_{eff}
T-optimal $\xi^{*T} = \begin{Bmatrix} 0.1730 & 1.131 & 5.104 & 28.08 \\ 0.0187 & 0.1216 & 0.2163 & 0.6434 \end{Bmatrix}$	0.2428	0.2333	1	1
D-optimal, model 1 $\xi^{*D_1} = \begin{Bmatrix} 0.2560 & 1 & 3.906 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$	1	0.7581	0	0.8162
D-optimal, model 2 $\xi^{*D_2} = \begin{Bmatrix} 0.2453 & 1.272 & 8.932 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$	0.8202	1	0	0.6565
D-optimal, product $\xi^{*P} = \begin{Bmatrix} 0.2578 & 1.143 & 6.412 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$	0.9279	0.9512	0	0.8574
$P\|T$-optimal $\xi^{*P\|T}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*T} & 0.2594 & 1.128 \\ 0.75\boldsymbol{w}^{*T} & 0.1683 & 0.0817 \end{Bmatrix}$	0.5541	0.5114	0.7824	0.9603
$\xi^{*P\|T}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*T} & 0.2590 & 1.141 & 6.910 \\ 0.5\boldsymbol{w}^{*T} & 0.2289 & 0.1710 & 0.1001 \end{Bmatrix}$	0.6770	0.6632	0.5254	0.9139
$\xi^{*P\|T}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*T} & 0.2584 & 1.143 & 6.578 \\ 0.25\boldsymbol{w}^{*T} & 0.2814 & 0.2525 & 0.2161 \end{Bmatrix}$	0.8014	0.8088	0.2634	0.8831
$T\|P$-optimal $\xi^{*T\|P}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*P} & 0.1730 & 1.131 & 5.104 & 28.08 \\ 0.25\boldsymbol{w}^{*P} & 0.0140 & 0.0912 & 0.1622 & 0.4826 \end{Bmatrix}$	0.4946	0.4746	0.7821	0.9644
$\xi^{*T\|P}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*P} & 0.1730 & 1.131 & 5.104 & 28.08 \\ 0.5\boldsymbol{w}^{*P} & 0.0093 & 0.0608 & 0.1081 & 0.3217 \end{Bmatrix}$	0.6597	0.6465	0.5251	0.9287
$\xi^{*T\|P}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*P} & 0.1730 & 1.131 & 5.104 & 28.08 \\ 0.75\boldsymbol{w}^{*P} & 0.0047 & 0.0304 & 0.0540 & 0.1609 \end{Bmatrix}$	0.8005	0.8025	0.2635	0.8931
Hybrid $\xi^{\text{hybrid}}_{\alpha=0.25} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.25\boldsymbol{w}^{*P} & 0.75\boldsymbol{w}^{*T} \end{Bmatrix}$	0.4946	0.4746	0.7821	0.9644
$\xi^{\text{hybrid}}_{\alpha=0.5} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.5\boldsymbol{w}^{*P} & 0.5\boldsymbol{w}^{*T} \end{Bmatrix}$	0.6597	0.6465	0.5251	0.9287
$\xi^{\text{hybrid}}_{\alpha=0.75} = \begin{Bmatrix} \boldsymbol{x}^{*P} & \boldsymbol{x}^{*T} \\ 0.75\boldsymbol{w}^{*P} & 0.25\boldsymbol{w}^{*T} \end{Bmatrix}$	0.8005	0.8025	0.2635	0.8931

Table 3.3: Efficiencies of near-optimal designs for two competing models for rise and decay.

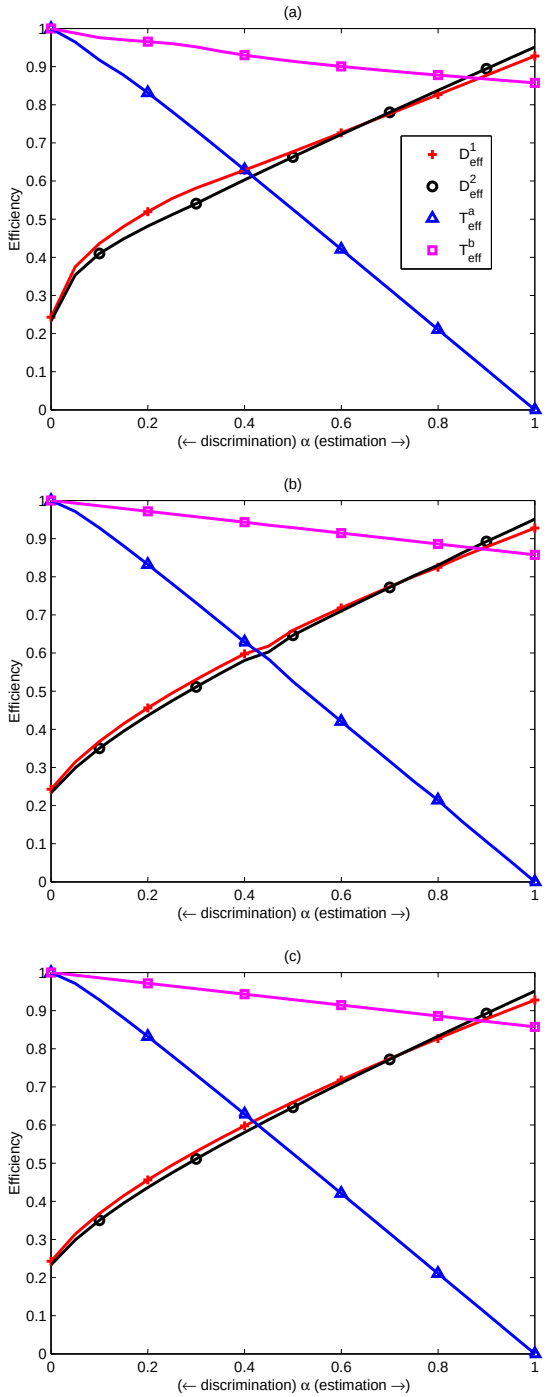

Figure 3.4: Efficiencies of designs for two models for rise and decay: (a) $P|T$-optimal designs; (b) $T|P$-optimal designs; (c) hybrid designs. In each case, α represents the importance placed on parameter estimation.

of fairly simple nonlinear models. The next chapter puts some of this work to use in the context of a considerably more complex pair of models incorporating multiple responses, random parameters, and ordinary differential equations with no analytic solution.

Chapter 4

Optimal design for nonlinear mixed effects models with multiple responses

The methods of optimal design for nonlinear models in the previous chapter will now be applied to a real-world example of a pharmacokinetic (PK) study, in which the aim is to choose a model which describes the time course of the changing concentration of a drug in the body, and to efficiently estimate the parameters in the model. As with many PK studies, the models involved are rather complex, involving random effects (to allow the PK parameters to vary between subjects) and multiple responses (we are interested in modelling the concentration of both the drug and its main metabolite). Optimal design for nonlinear mixed effects models has been given some attention in the literature recently, particularly in regards to PK experiments (Mentré *et al.*, 1997; Retout and Mentré, 2003; Green and Duffull, 2003). The use of multiple responses, however, introduces some theoretical challenges and also adds computational complexity to an already formidable problem.

4.1 Introduction

In this chapter, the problem of optimal design for nonlinear models with multiple responses is addressed through a real example in which an experiment is designed for the study of the PK of the drug itraconazole. Part of the aim of the study is to discriminate between two candidate models, hence some ideas of the previous chapter are employed here. The

inclusion of multiple responses in the models motivates the investigation of techniques for incorporating all responses into the design process.

4.1.1 Example: Background and significance

Allergic bronchopulmonary aspergillosis is a complication of cystic fibrosis (CF), occurring in approximately 10% of patients mostly after the age of 6 years (Moss, 2002). A study has shown that patients achieved an improvement in the inflammatory response when itraconazole was given in combination with systemic glucocorticosteroids (Stevens *et al.*, 2000).

Although relatively little is presently known about the pharmacokinetics of itraconazole in patients with CF, there is evidence which suggests that it is different to the PK of the drug in non-CF patients (Hedman *et al.*, 1988; Jusko *et al.*, 1975; Finkelstein and Hall, 1979). It is also known that there is often large variability in the pharmacokinetic parameters (eg. absorption rate constant, volume of distribution, and clearance) of drugs administered to CF patients (Reed, 1997; Spino, 1991). It has been observed that high doses of itraconazole are required in CF patients for the treatment of allergic bronchopulmonary aspergillosis (Skov *et al.*, 2002).

Itraconazole is extensively metabolised by the hepatic cytochrome P450 isoenzymes and has at least 30 metabolites, each representing less than 1%–5% of the dose. The main metabolite is considered to be hydroxyitraconazole, which also exhibits antifungal activity against a variety of species (Heykants *et al.*, 1989; Hostetler *et al.*, 1993; Van Cutsem, 1989).

4.1.2 Structural models

The pharmacokinetics of itraconazole and its main metabolite has been shown to exhibit the behaviour of two linked compartmental models. Figure 4.1 shows the general structure of the models, including the rates of mass transfer between each compartment. The amounts of itraconazole in the gut, central and peripheral compartments are given by A_1, A_2 and A_3, respectively; the amount of its main metabolite (hydroxyitraconazole) is denoted by A_4. The elimination from the parent compartment to a compound other than hydroxyitraconazole (shown in Figure 4.1) has been described by some as linear (elimination rate $F_{20}CL_2/V_2$)

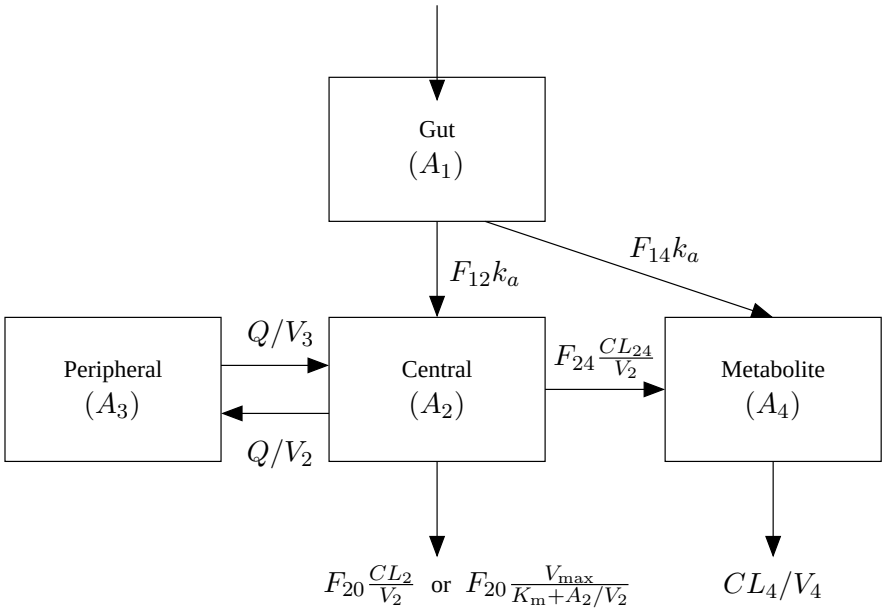

Figure 4.1: Compartmental model of the pharmacokinetics of itraconazole and hydroxyitraconazole.

(Koks *et al.*, 2002), and by others as nonlinear (elimination rate $F_{20}V_{\max}/(K_m + A_2/V_2)$) (Barone *et al.*, 1993; Van Peer *et al.*, 1989; Heykants *et al.*, 1989; Hardin *et al.*, 1988).

The differential equations describing the two models (linear and nonlinear) from Figure 4.1 are shown in Equations (4.1–4.8).

Model 1. The linear model:

$$\frac{\mathrm{d}A_1}{\mathrm{d}t} = -F_{12}k_aA_1 - F_{14}k_aA_1 \tag{4.1}$$

$$\frac{\mathrm{d}A_2}{\mathrm{d}t} = F_{12}k_aA_1 + \frac{Q}{V_3}A_3 - \frac{Q}{V_2}A_2 - F_{24}\frac{CL_{24}}{V_2}A_2 - F_{20}\frac{CL_2}{V_2}A_2 \tag{4.2}$$

$$\frac{\mathrm{d}A_3}{\mathrm{d}t} = \frac{Q}{V_2}A_2 - \frac{Q}{V_3}A_3 \tag{4.3}$$

$$\frac{\mathrm{d}A_4}{\mathrm{d}t} = F_{14}k_aA_1 + F_{24}\frac{CL_{24}}{V_2}A_2 - \frac{CL_4}{V_4}A_4 \tag{4.4}$$

Model 2. The nonlinear model:

$$\frac{\mathrm{d}A_1}{\mathrm{d}t} = -F_{12}k_aA_1 - F_{14}k_aA_1 \tag{4.5}$$

$$\frac{\mathrm{d}A_2}{\mathrm{d}t} = F_{12}k_aA_1 + \frac{Q}{V_3}A_3 - \frac{Q}{V_2}A_2 - F_{24}\frac{CL_{24}}{V_2}A_2 - F_{20}\frac{V_{\max}}{K_{\mathrm{m}} + A_2/V_2}A_2 \tag{4.6}$$

$$\frac{\mathrm{d}A_3}{\mathrm{d}t} = \frac{Q}{V_2}A_2 - \frac{Q}{V_3}A_3 \tag{4.7}$$

$$\frac{\mathrm{d}A_4}{\mathrm{d}t} = F_{14}k_aA_1 + F_{24}\frac{CL_{24}}{V_2}A_2 - \frac{CL_4}{V_4}A_4 \tag{4.8}$$

Note that since the differential equations describe the rates of change of the *amount* of drug in each compartment (rather than the concentration), the last term in Equation (4.6) describing the Michaelis-Menten elimination is slightly different to the form commonly used. The divisor still involves the concentration (i.e. A_2/V_2), but we multiply by A_2 to give the rate of change in the amount. The units of $V_{\max}$ and K_{m} are still of the form 'mass/(volume $\times$ time)' and 'mass/volume', respectively. In both sets of differential equations, the initial conditions are

$$A_1(0) = \text{dose}, \quad A_2(0) = A_3(0) = A_4(0) = 0.$$

4.2 Theory

A model must be formed to describe the behaviour over time of the amount of the parent drug and its metabolite (A_2 and A_4), and hence the behaviour of their concentrations ($C_2 = A_2/V_2$ and $C_4 = A_4/V_4$).

It is assumed that the fractions F_{12}, F_{14}, F_{24} and F_{20} are known fixed constants (where F_{ij} is the fraction of the drug delivered from compartment i to compartment j for $j \neq 0$, and F_{20} is the fraction of the drug eliminated from compartment 2) and that their values are available from previous experiments. The parameters to be estimated which are common to both models are therefore the volumes of distribution (the amount of drug in the compartment divided by the concentration in the blood) V_2, V_3 and V_4; and the clearances (the volume of plasma from which the drug is completely removed per unit time) CL_4, CL_{24} and Q. The linear model has one additional clearance parameter, CL_2, while the nonlinear model also

has parameters $V_{\max}$ and K_{m}. Considering now that interest lies in the pharmacokinetics of itraconazole in both capsule and solution form, there are actually four models under consideration: the linear model for capsules, the linear model for solutions, the nonlinear model for capsules and the nonlinear model for solutions. The models for capsules and solutions have the same structure, they only differ in terms of their parameter values. Hence there are four sets of parameter values:

$$\boldsymbol{\beta}^{1,c} = (k_a^{1,c}, V_2^{1,c}, V_3^{1,c}, V_4^{1,c}, CL_4^{1,c}, CL_{24}^{1,c}, Q^{1,c}, CL_2^{1,c})', \tag{4.9}$$

$$\boldsymbol{\beta}^{2,c} = (k_a^{2,c}, V_2^{2,c}, V_3^{2,c}, V_4^{2,c}, CL_4^{2,c}, CL_{24}^{2,c}, Q^{2,c}, V_{\max}^{2,c}, K_{\mathrm{m}}^{2,c})', \tag{4.10}$$

$$\boldsymbol{\beta}^{1,s} = (k_a^{1,s}, V_2^{1,s}, V_3^{1,s}, V_4^{1,s}, CL_4^{1,s}, CL_{24}^{1,s}, Q^{1,s}, CL_2^{1,s})', \tag{4.11}$$

$$\boldsymbol{\beta}^{2,s} = (k_a^{2,s}, V_2^{2,s}, V_3^{2,s}, V_4^{2,s}, CL_4^{2,s}, CL_{24}^{2,s}, Q^{2,s}, V_{\max}^{2,s}, K_{\mathrm{m}}^{2,s})', \tag{4.12}$$

where for $\boldsymbol{\beta}^{u,f}$ the superscripts u and f indicate the structural model u ($u = 1$ for linear, $u = 2$ for nonlinear) and the form of dose f ($f = c$ for capsule, $f = s$ for solution).

For the time being, suppose that only one structural model and only one form of dose is under consideration. The superscripts u and f are dropped, and the p-vector of parameters is denoted by $\boldsymbol{\beta}$. Let $\boldsymbol{\theta}_i$ be the vector of parameter values for the i^{th} individual. Then for a vector of n_i sampling times $\boldsymbol{\xi}_i = (t_{i1}, \ldots, t_{in_i})'$, suppose that the n_i-vector of observations is modelled by

$$\boldsymbol{y}_i = \eta(\boldsymbol{\theta}_i, \boldsymbol{\xi}_i) + \boldsymbol{\epsilon}_i.$$

Employing a nonlinear mixed effects model, it is assumed that each of the parameters varies randomly between individuals. Let $\boldsymbol{\beta}$ be the vector of fixed effects and let $\boldsymbol{b}_i$ be the vector of random effects for individual i. So the random effects are written as either $\boldsymbol{\theta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i$ (for additive random effects) or $\boldsymbol{\theta}_i = \boldsymbol{\beta}\exp(\boldsymbol{b}_i)$ (for exponential random effects). Thus the structural model can be written in an alternative form, $\eta(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\xi}_i)$. It is assumed that $\boldsymbol{b}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_p)$ and $\boldsymbol{0}$ is the vector of zeros. It is also assumed that, as in Retout and Mentré (2003), the variances of the normally distributed error terms $\boldsymbol{\epsilon}_i$ (conditional on the random effects $\boldsymbol{b}_i$) have additive and proportional components, i.e. $\boldsymbol{\epsilon}_i|\boldsymbol{b}_i \sim \mathrm{N}(0, \mathrm{diag}(\sigma_A + \sigma_P\eta(\boldsymbol{\theta}_i, \boldsymbol{\xi}_i))^2$ for some σ_A and σ_P. The vector of all variance parameters is $\boldsymbol{\lambda} = (\omega_1, \ldots, \omega_p, \sigma_A, \sigma_P)'$ Denote the entire vector of model parameters by $\boldsymbol{\Psi} = (\boldsymbol{\beta}', \boldsymbol{\lambda}')'$.

4.2.1 Optimal design

Model discrimination aside (momentarily), the objective of the experiment is to obtain estimates of the relevant parameters with the greatest possible efficiency. For this purpose, the Fisher information matrix is employed.

First some slightly different notation is introduced for use in this chapter. For consistency with the existing literature on optimal design for PK models, an experimental design (the 'population design' in the PK literature) with N independent subjects is redefined as $\Xi = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N)'$, where $\boldsymbol{\xi}_i$ is the 'elementary design' (eg. set of sampling times) for individual i. The population design is usually limited to Q groups, with N_q subjects allocated to the qth group:

$$\Xi = \begin{Bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \ldots & \boldsymbol{\xi}_Q \\ N_1 & N_2 & \ldots & N_Q \end{Bmatrix}, \tag{4.13}$$

with $\sum_{q=1}^{Q} N_q = N$.

Recall from Section 2.3.1 that the Fisher information matrix is given by

$$\boldsymbol{M}(\boldsymbol{\Psi}, \Xi) = \mathrm{E}\left[-\frac{\partial^2 \ell(\boldsymbol{\Psi}; \boldsymbol{y})}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'}\right],$$

where $\ell(\boldsymbol{\Psi}; \boldsymbol{y})$ is the log-likelihood of the vector of observations $\boldsymbol{y} = (y_1, \ldots, y_N)'$ for the population parameters Ψ. This may be expressed as the weighted sum of the information matrices for the Q elementary designs:

$$\boldsymbol{M}(\boldsymbol{\Psi}, \Xi) = \sum_{q=1}^{Q} N_q \boldsymbol{M}(\boldsymbol{\Psi}, \boldsymbol{\xi}_q).$$

Maximising the determinant of $\boldsymbol{M}(\boldsymbol{\Psi}, \boldsymbol{\Xi})$ leads to the D-optimal design, as described in Section 2.3.2.

Unfortunately, for nonlinear mixed effects models, the information matrix cannot be written down in closed form. For an elementary design $\boldsymbol{\xi}$, Mentré *et al.* (1997) have shown that by using a first-order Taylor series expansion of $\eta(\boldsymbol{\theta}, \boldsymbol{\xi})$ around the expectation of the random effects, an approximation to the Fisher information matrix for an elementary design $\boldsymbol{\xi}$ is given by

$$\boldsymbol{M}(\boldsymbol{\Psi}, \boldsymbol{\xi}) \approx \frac{1}{2} \begin{bmatrix} \boldsymbol{A}(\boldsymbol{E}, \boldsymbol{S}) & \boldsymbol{C}(\boldsymbol{E}, \boldsymbol{S}) \\ \boldsymbol{C}'(\boldsymbol{E}, \boldsymbol{S}) & \boldsymbol{B}(\boldsymbol{E}, \boldsymbol{S}) \end{bmatrix}, \tag{4.14}$$

where

$$(\boldsymbol{A}(\boldsymbol{E},\boldsymbol{S}))_{mn} = 2\frac{\partial \boldsymbol{E}'}{\partial \beta_m}\boldsymbol{S}^{-1}\frac{\partial \boldsymbol{E}}{\partial \beta_n}, \quad m,n = 1,\ldots,p \tag{4.15}$$

$$(\boldsymbol{B}(\boldsymbol{E},\boldsymbol{S}))_{mn} = \operatorname{tr}\left(\frac{\partial \boldsymbol{S}}{\partial \lambda_m}\boldsymbol{S}^{-1}\frac{\partial \boldsymbol{S}}{\partial \lambda_n}\boldsymbol{S}^{-1}\right) \quad m,n = 1,\ldots,p+2 \tag{4.16}$$

$$(\boldsymbol{C}(\boldsymbol{E},\boldsymbol{S}))_{mn} = \operatorname{tr}\left(\frac{\partial \boldsymbol{S}}{\partial \lambda_m}\boldsymbol{S}^{-1}\frac{\partial \boldsymbol{S}}{\partial \beta_n}\boldsymbol{S}^{-1}\right) \quad m = 1,\ldots,p+2,\ n = 1,\ldots,p \tag{4.17}$$

Here $\boldsymbol{E}$ and $\boldsymbol{S}$ represent approximations to the expectation and variance of the observations, respectively:

$$\mathrm{E}(\boldsymbol{y}) \approx \boldsymbol{E} = \eta(\boldsymbol{\beta},\boldsymbol{0},\boldsymbol{\xi}) \tag{4.18}$$

$$\mathrm{Var}(\boldsymbol{y}) \approx \boldsymbol{S} = \left[\frac{\partial \eta(\boldsymbol{\beta},\boldsymbol{0},\boldsymbol{\xi})}{\partial \boldsymbol{b}'}\right]\boldsymbol{\Omega}\left[\frac{\partial \eta(\boldsymbol{\beta},\boldsymbol{0},\boldsymbol{\xi})}{\partial \boldsymbol{b}'}\right]' + \operatorname{diag}\left(\sigma_A + \sigma_P \eta(\boldsymbol{\beta},\boldsymbol{0},\boldsymbol{\xi})\right)^2 \tag{4.19}$$

For computational efficiency, it is assumed that $\boldsymbol{S}$ is independent of $\boldsymbol{\beta}$, as in Retout *et al.* (2002), which makes $\boldsymbol{C}(\boldsymbol{E},\boldsymbol{S})$ a matrix of zeros. From the simulations carried out in Section 4.4.2, the impact of this assumption on the final design is found to be minimal. This is a worthwhile assumption, as it will be seen that this is a computationally intensive optimisation, and any justifiable simplification is welcome.

4.2.2 Parameter estimation in multiresponse situations

Consider for a moment that the observations are in fact modelled by a simpler fixed effects model with multiple response types. Suppose that the outcomes of an experiment on a single subject with r simultaneous responses can be modelled as

$$y_i^{(a)} = \eta^{(a)}(\boldsymbol{\theta}, t_i) + \epsilon_i^{(a)} \quad (a = 1,\ldots,r;\ i = 1,\ldots,n),$$

where the $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_p)'$ is now a p-vector of fixed effects parameters; $\mathrm{E}(\epsilon_i^{(a)}) = 0$; $\mathrm{E}(\epsilon_i^{(a)}\epsilon_j^{(b)}) = 0$ for $i \neq j$; $\mathrm{E}((\epsilon_i^{(a)})^2) = \sigma_a^2 = \sigma_{aa}$; and $\mathrm{E}(\epsilon_i^{(a)}\epsilon_i^{(b)}) = \rho_{ab}\sigma_a\sigma_b = \sigma_{ab}$ for $a \neq b$. So

the covariance matrix of responses for the i^{th} sampling time is given by

$$\boldsymbol{\Sigma} = \{\sigma_{ab}\}_{a,b=1,\ldots,r} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1r} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{r1} & \sigma_{r2} & \cdots & \sigma_{rr} \end{bmatrix}$$

where $\sigma_{ab} = \sigma_{ba}$. Let $\boldsymbol{\Sigma}^{-1} = \{\sigma^{ab}\}_{a,b=1,\ldots,r}$.

Again, denote the elementary design by $\boldsymbol{\xi} = (t_1, \ldots, t_n)'$. Draper and Hunter (1966) have shown that the information matrix for multiple responses is given by

$$\boldsymbol{M}(\boldsymbol{\theta}, \boldsymbol{\xi}) = \sum_{a=1}^{r} \sum_{b=1}^{r} \sigma^{ab} \boldsymbol{F}_a' \boldsymbol{F}_b, \tag{4.20}$$

where

$$\boldsymbol{F}_a = \left[\frac{\partial \eta_u^{(a)}(\boldsymbol{\theta}, t_1)}{\partial \boldsymbol{\theta}}, \cdots, \frac{\partial \eta_u^{(a)}(\boldsymbol{\theta}, t_n)}{\partial \boldsymbol{\theta}} \right], \quad a = 1, \ldots, r.$$

The D-optimal design then maximises the determinant of $\boldsymbol{M}(\boldsymbol{\theta}, \boldsymbol{\xi})$ as given in Equation (4.20).

It follows that if the information matrix in Equation (4.20) is simplified to

$$\boldsymbol{M}(\boldsymbol{\theta}, \boldsymbol{\xi}) = \sum_{a=1}^{r} \sum_{b=1}^{r} \boldsymbol{F}_a' \boldsymbol{F}_b,$$

i.e. if it assumed that the σ^{ab} are all equal, then this is equivalent to maximising $|\boldsymbol{F}'\boldsymbol{F}|$, where

$$\begin{aligned} \boldsymbol{F} &= \sum_{a=1}^{r} \boldsymbol{F}_a \\ &= \left[\frac{\partial \sum_{a=1}^{r} \eta^{(a)}(\boldsymbol{\theta}, t_1)}{\partial \boldsymbol{\theta}}, \cdots, \frac{\partial \sum_{a=1}^{r} \eta^{(a)}(\boldsymbol{\theta}, t_n)}{\partial \boldsymbol{\theta}} \right]. \end{aligned}$$

This will greatly simplify computation, as all that is required for the criterion is to sum the r responses in the model, and then calculate the Fisher information matrix as in the single response case. While this assumption may not be realistic for this study, there appear to be no published values for the covariance of itraconazole and hydroxyitraconazole, so this algebraically convenient (and hence computationally simple) simplification is reasonable in these circumstances. Again, the extensive simulation studies in Section 4.4.2 demonstrate that this assumption has little impact on the final design.

This idea is extended to the more complex nonlinear mixed effects models with heteroscedastic error. To accomplish this, the Fisher information matrix described in Equations (4.14–4.19) is computed, replacing $\eta(\boldsymbol{\beta}, 0, \boldsymbol{\xi})$ with $\sum_{a=1}^{r} \eta^{(a)}(\boldsymbol{\beta}, 0, \boldsymbol{\xi})$. In the itraconazole example, the only responses of interest are the concentrations in the 2nd and 4th compartments. Hence the sum is only over $a = 2, 4$.

4.2.3 Model discrimination

There are two competing structural models to describe the pharmacokinetics of itraconazole. The first involves linear elimination of the parent drug, the second involves a nonlinear Michaelis-Menten process. The T-optimality criterion for model discrimination, described in Section 2.3.3, could be applicable here. However, T-optimality is quite expensive computationally (which is a major concern in this complex scenario), and is not efficient for parameter estimation (see Section 3.2 for some examples of this). A viable alternative is to use the product of the determinants of the information matrices, raised to the power of the reciprocal of the number of population parameters (or equivalently, the product of efficiencies, as suggested by Atkinson and Cox (1974)). We have also seen from Section 3.2 suggestions that in some circumstances this method can produce designs which are useful for both goals: parameter estimation and model discrimination. This was evaluated by the simulations in Section 4.4.2.

4.2.4 Combining capsule and solution responses

As mentioned in Section 4.1.2, there are four models under consideration. However, only the linear and nonlinear structural models must be discriminated between. The structure of the models for capsules and solutions only differ in the values of the parameters, not the structure of the models. To do this, a capsule dose is administered, followed by a solution dose at a time, determined by the clinical staff, after which it is assumed that there is no residual amount of the capsule dose. It was also assumed by the clinical staff that there is no effect of the order of the doses. An elementary design $\boldsymbol{\xi}$ includes sampling times for the capsule ($\boldsymbol{\xi}^c$) and solution ($\boldsymbol{\xi}^s$). The responses of subject i to both doses are combined to

give the overall response. So the parameters for capsules and solutions for a given structural model must also be combined, i.e. for $u = 1$ and 2 (linear and nonlinear models, respectively), $\boldsymbol{\theta}^u = ((\boldsymbol{\theta}^{u,c})', (\boldsymbol{\theta}^{u,s})')'$. In the context of nonlinear mixed effects models with additive random effects, we now have the vector of parameters in model u for subject i given by

$$\begin{aligned} \boldsymbol{\theta}_i^u &= ((\boldsymbol{\theta}_i^{u,c})', (\boldsymbol{\theta}_i^{u,s})')' \\ &= ((\boldsymbol{\beta}^{u,c} + \boldsymbol{b}_i^{u,c})', (\boldsymbol{\beta}^{u,s} + \boldsymbol{b}_i^{u,s})')' \qquad (4.21) \\ &= \boldsymbol{\beta}^u + \boldsymbol{b}_i^u \qquad (4.22) \end{aligned}$$

where $\boldsymbol{\beta}^u = ((\boldsymbol{\beta}^{u,c})', (\boldsymbol{\beta}^{u,s})')'$ and $\boldsymbol{b}_i^u = ((\boldsymbol{b}_i^{u,c})', (\boldsymbol{b}_i^{u,s})')'$, with the random effects for dose form f ($f = c, s$) for each subject being normally distributed with zero mean and with covariance matrix $\boldsymbol{\Omega}^{u,f} = \text{diag}(\omega_1^{u,f}, \ldots, \omega_{p^{u,f}}^{u,f})$. The combined covariance matrix for model u is then $\boldsymbol{\Omega}^u = \text{diag}(\boldsymbol{\Omega}^{u,c}, \boldsymbol{\Omega}^{u,s})$. Equations (4.21) and (4.22) may be adjusted for exponential random effects by replacing $\boldsymbol{\beta} + \boldsymbol{b}_i$ with $\boldsymbol{\beta}\exp(\boldsymbol{b}_i)$ (for any superscripts on $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$).

The entire vector of population parameters under model u is then

$$\boldsymbol{\Psi}^u = ((\boldsymbol{\Psi}^{u,c})', (\boldsymbol{\Psi}^{u,s})')', \qquad (4.23)$$

with

$$\boldsymbol{\Psi}^{u,f} = ((\boldsymbol{\beta}^{u,f})', \omega_1^{u,f}, \ldots, \omega_{p^{u,f}}^{u,f}, \sigma_A^{u,f}, \sigma_P^{u,f})'.$$

In this notation, the combined responses to the capsule and solution doses for compartment a (where $a = 2$ or 4) can be written

$$\eta_u^{(a)}(\boldsymbol{\beta}^u, \boldsymbol{b}_i^u, \boldsymbol{\xi}_i) = (\eta_u^{(a)}(\boldsymbol{\beta}^{u,c}, \boldsymbol{b}_i^{u,c}, \boldsymbol{\xi}_i^c)', \eta_u^{(a)}(\boldsymbol{\beta}^{u,s}, \boldsymbol{b}_i^{u,s}, \boldsymbol{\xi}_i^s)')'.$$

4.2.5 Final criterion

Hence the overall product criterion to be maximised is given by

$$D_P(\boldsymbol{\Psi}^1, \boldsymbol{\Psi}^2, \Xi) = |\boldsymbol{M}^1(\boldsymbol{\Psi}^1, \Xi)|^{1/p_1} |\boldsymbol{M}^2(\boldsymbol{\Psi}^2, \Xi)|^{1/p_2}, \qquad (4.24)$$

where

$$\boldsymbol{M}^u(\boldsymbol{\Psi}^u, \Xi) = \sum_{q=1}^{Q} N_q \boldsymbol{M}^u(\boldsymbol{\Psi}^u, \boldsymbol{\xi}_q), \quad u = 1, 2.$$

$\boldsymbol{\Psi}^u$ is the vector of population parameters for the u^{th} model. For the u^{th} model, the elementary information matrix is given by

$$\boldsymbol{M}^u(\boldsymbol{\Psi}^u, \boldsymbol{\xi}_q) \approx \frac{1}{2} \begin{bmatrix} \boldsymbol{A}(\boldsymbol{E}^u, \boldsymbol{S}^u) & \boldsymbol{C}(\boldsymbol{E}^u, \boldsymbol{S}^u) \\ \boldsymbol{C}'(\boldsymbol{E}^u, \boldsymbol{S}^u) & \boldsymbol{B}(\boldsymbol{E}^u, \boldsymbol{S}^u) \end{bmatrix},$$

with $\boldsymbol{A}(\cdot,\cdot)$, $\boldsymbol{B}(\cdot,\cdot)$ and $\boldsymbol{C}(\cdot,\cdot)$ as described in Equations (4.15-4.16), and

$$\mathrm{E}(\boldsymbol{y}) \approx \boldsymbol{E}^u = \sum_{a=2,4} \eta_u^{(a)}(\boldsymbol{\beta}^u, 0, \boldsymbol{\xi}_q) \tag{4.25}$$

$$\begin{aligned} \mathrm{Var}(\boldsymbol{y}) \approx \boldsymbol{S}^u = & \left[\frac{\partial \sum_{a=2,4} \eta_u^{(a)}(\boldsymbol{\beta}^u, 0, \boldsymbol{\xi}_q)}{\partial \boldsymbol{b}'} \right] \boldsymbol{\Omega}^u \left[\frac{\partial \sum_{a=2,4} \eta_u^{(a)}(\boldsymbol{\beta}^u, 0, \boldsymbol{\xi}_q)}{\partial \boldsymbol{b}'} \right]' \\ & + \mathrm{diag} \left(\sigma_a + \sigma_p \sum_{a=2,4} \eta_u^{(a)}(\boldsymbol{\beta}^u, 0, \boldsymbol{\xi}_q) \right)^2, \end{aligned} \tag{4.26}$$

where the sums are over the responses for compartments 2 and 4.

4.3 Methods

In this section a description is given of the constraints of the sampling design, procedures used in selecting appropriate parameter and constant values, a number of simplifications to the models (to fit design constraints), and a brief overview of computational methods used to optimise the design criterion.

4.3.1 Design constraints

A number of limitations on the experimental design were imposed by the clinical staff for reasons of time and expense. A maximum of thirty patients were to be enrolled and studied on two occasions. The number of blood samples was also constrained to be four samples per occasion. A single 200 mg dose of the drug in capsules was to be given to each patient on the first occasion, followed by a single 200 mg dose in solution on the second occasion.

A time constraint of approximately two months was placed on the design of the experiment, from exploration of the literature to computation of the final design.

4.3.2 Patients

Thirty adults with CF who are admitted to hospital for treatment of a chest exacerbation will receive two doses of itraconazole (one 200 mg dose in capsules and one 200 mg dose as an oral solution) for this study.

The participating patients are divided randomly into three groups, with each group being assigned one of three elementary sampling designs.

Accordingly patients will be required to take two *Sporonax® Capsules* as one dose (200 mg itraconazole), which is a standard adult dose. Four blood samples (0.5 mL) will be then drawn by finger prick according to the optimum sampling times of the design.

After a wash-out period of at least three days (after the initial dose), the same dose of 200 mg (20 mL of 100 mg/10 mL solution) itraconazole would be taken as *Sporonax® Oral Solution*. Four blood samples (0.5 mL) will again be drawn by finger prick according to the optimum sampling times of the design.

4.3.3 Simplification of models

The design optimisation problem under consideration is quite computationally intensive. Any justifiable reduction in the number of parameters to be estimated (which leads to a reduction in the dimension of the Fisher information matrix) is greatly beneficial.

It is a fundamental belief in pharmacology that the different forms of dose will affect the input process only. So rather than treat the model parameters for capsule and solution doses as being entirely separate, as in Equations (4.9–4.12), it is proposed that there are only two sets of parameters to be estimated, the first for the linear model and the second for the nonlinear model:

$$\boldsymbol{\beta}^1 = (k_a^1, V_2^1, V_3^1, V_4^1, CL_4^1, CL_{24}^1, Q^1, CL_2^1)', \tag{4.27}$$

$$\boldsymbol{\beta}^2 = (k_a^2, V_2^2, V_3^2, V_4^2, CL_4^2, CL_{24}^2, Q^2, V_{\max}^2, K_{\mathrm{m}}^2)'. \tag{4.28}$$

The parameters k_a^u represent the absorption rate for capsules (i.e. they are strictly $k_a^{u,c}$). The distinction between absorption rates is instead addressed by a multiplicative factor $F_{k_a} = 0.9452/0.4159 \approx 2.273$, the ratio of the absorption rate for solutions and the absorption

rate for capsules (not to be estimated). So the response for the solution under model u is given by using the absorption rate $F_{k_a} k_a^u$. The smaller of the two absorption rates is used for this design, as it will be the most difficult to estimate.

In addition to the computational intensity of the problem, the design constraints (outlined in Section 4.3.1) mean that not all of the parameters (of which there is still a large number) will be estimable. Preliminary investigation of expected standard errors was carried out by calculating the information matrix for an empirical 'rich' sampling design, in which 14 samples are taken from each patient over almost the entire dosing interval (72 hours) with sampling times given by (0.01, 0.5, 1, 1.5, 2, 4, 6, 8, 12, 18, 24, 36, 48, 70). The expected standard errors are given by the square roots of the diagonal elements of the inverse of the information matrix. This investigation showed that under both linear and nonlinear models, the parameters CL_{24}, Q and V_3 will have very imprecise estimates under the constraints of the sampling design. Further to this, K_{m} is not estimable with any precision under the nonlinear model. These four parameters are hence treated as fixed, and are not considered in the design process.

To further simplify computation, the residual amounts of drug and metabolite in each compartment at the time of the second dose were assumed to be insignificant (i.e. there is no carryover effect). This was found to have a negligible effect on the criterion, which was calculated for the design mentioned above, both including and ignoring the residual dose.

4.3.4 Parameter and constant values

Finding the D-optimal design for any model which is nonlinear in its parameters requires prior specification of the model parameters. Such designs are known as 'locally' D-optimal.

The parameter values (and values of other constants) used in the construction of the design were taken from a variety of sources. To find the local optimum design, parameters which are common to both linear and nonlinear elimination models (eg. V_2, CL_4) are given the same value for both model types, despite the treatment of the parameters as separate entities when constructing the design criterion. For example, V_2^1 and V_2^2 are not the same population parameters, but for convenience the same value is assigned to both parameters

in the construction of the optimum design as only one value was available in the literature.

A recently published population pharmacokinetic study by Koks *et al.* (2003) gives most of the parameter values needed. In particular, it gives the clearance of itraconazole from the central compartment and its volume of distribution in the central compartment, and therefore the elimination rate constant from the central compartment (k_{20}), and also the distribution rate constants (k_{23}, k_{32}) and the formation rate constant (k_{24}).

These authors did not report any nonlinear elimination of the drug. But in several other pharmacokinetic studies, nonlinear pharmacokinetic behaviour was shown for itraconazole (Heykants *et al.*, 1989; Barone *et al.*, 1993; Van de Velde *et al.*, 1996; Van Peer *et al.*, 1989). Information about the Michaelis-Menten kinetic parameters (K_{m} and $V_{\max}$) was consequently collected from other sources (Barone *et al.*, 1993).

Since the aim is to study the relative bioavailability of the oral solution compared to the capsule formulation of itraconazole in patients with CF, the different absorption rate constants (k_a^s and k_a^c) must be estimated as well. This was done using

$$k_a = \frac{\log_e 2}{t_{1/2,\,\mathrm{absorption}}},$$

where $t_{1/2,\,\mathrm{absorption}}$ is estimated by $t_{\max}/3$. The times of peak plasma concentration for itraconazole in the oral solution and the capsule formulation from the drug information web page of the Food and Drug Administration (FDA) (Janssen Pharmaceutica Products, L.P., 2002b) were used.

As no further information on PK parameters was available, the remaining model parameters were estimated by 'best guess'. Table 4.1 shows, along with all previously mentioned parameters and constants, the values chosen for the elimination rate constant, the volume of distribution, and the clearance of the metabolite (k_{40}, V_4, and CL_4, respectively). The best guess values of these parameters were chosen as those that produced a simulated concentration-time curve that was closely comparable to profiles provided by the authors of other clinical pharmacokinetic studies (Heykants *et al.*, 1989; Barone *et al.*, 1998). Deterministic simulations were performed using MATLAB (these are shown in Figure 4.2).

Residual variability was assumed to have both additive and proportional components, with $\sigma_A = 5 \times 10^{-3}$ mg/L (half of the lowest concentration detectible by the assay) and

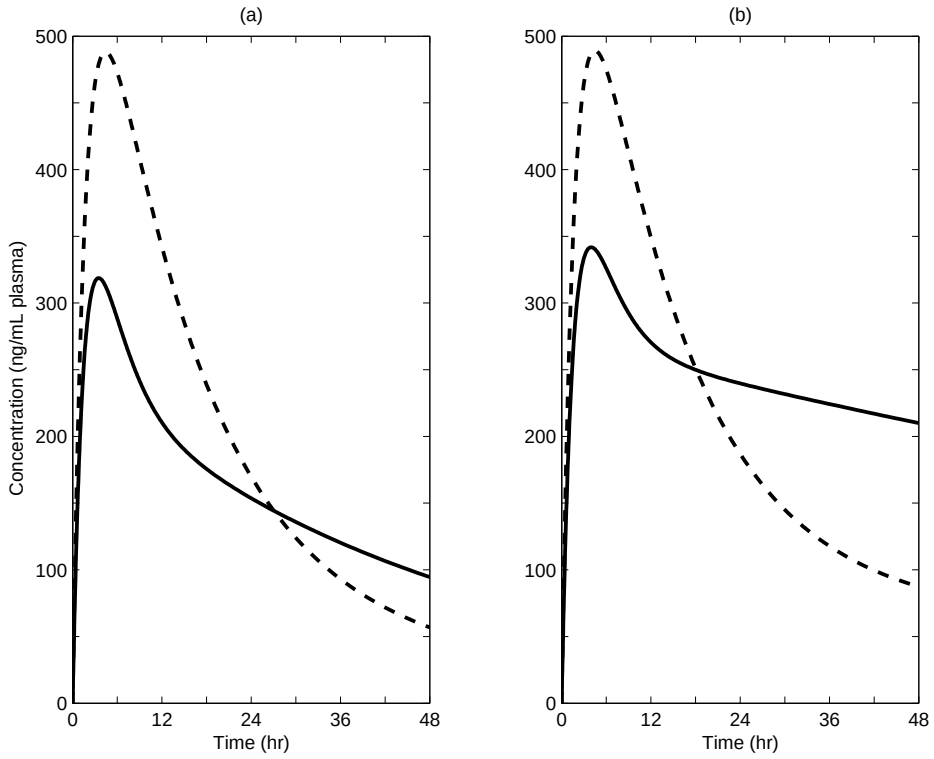

Figure 4.2: Concentration-time profiles of itraconazole (solid line) and hydroxyitraconazole (dashed line) after a 200 mg solution dose of itraconazole, with (a) linear elimination of the drug; and (b) nonlinear elimination of the drug. The values of the various parameters and constants used to produce the plots are given in Table 4.1.

$\sigma_P = 0.15$.

Between subject variability was assumed to be lognormal with variance 0.1, i.e. $\boldsymbol{\theta}_i = \boldsymbol{\beta}\exp(\boldsymbol{b}_i)$ with $\boldsymbol{b}_i \sim N(0, 0.1\boldsymbol{I}_p)$ for all parameters, where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix.

4.3.5 Computational methods

A set of MATLAB routines, POPT® Version 2.0, was used to calculate the optimal design, optimising the criterion given in Section 4.2.5. This software allows the user to find D-optimal sampling times by specifying the number of blood samples, the upper and lower boundaries for the samples, the dose interval for taking blood samples, and the number of groups in the study. Duffull *et al.* (2005) gives further description of the software used. POPT® was modified to incorporate the product criterion and numerical solutions to the

Table 4.1: Parameter and constant values used to calculated the optimal design. References: (A) = Janssen Pharmaceutica Products, L.P. (2002b); (B) = Koks *et al.* (2003); (C) = Barone *et al.* (1993); (D) = Janssen Pharmaceutica Products, L.P. (2002a); (E) = best guess derived from MATLAB simulation.

Variable	Value	Source
k_a^s = absorption rate constant, solution	0.945 h^{-1}	(A)
k_a^c = absorption rate constant, capsule	0.416 h^{-1}	(A)
F_{k_a} = absorption rate multiplicative factor	2.27	$= k_a^s/k_a^c$
k_{20} = elimination rate constant from the central compartment, linear model	7.6×10^{-2} h^{-1}	(B)
k_{23} = distribution rate constant	0.126 h^{-1}	(B)
k_{32} = distribution rate constant	0.15 h^{-1}	(B)
k_{24} = formation rate constant	2.93×10^{-3} h^{-1}	(B)
k_{40} = elimination rate constant for hydroxyitraconazole	7.6×10^{-2} h^{-1}	$= k_{20}$
CL_2 = clearance of itraconazole	27.9 L/h	(B)
CL_4 = clearance of the metabolite	1.75 L/h	$= k_{40}V_4$
Q = inter-compartmental clearance	46 L/h	$= k_{23}V_2$
$V_{\max}$ = theoretical maximum rate of the process	9.54 ng/(mL·h)	(C)
K_{m} = Michaelis-Menten constant	329 ng/mL	(C)
V_2 = volume of central compartment	365 L	(B)
V_4 = volume of metabolite compartment	23 L	$\approx \frac{1}{16}V_2$
V_3 = volume in peripheral compartment	307 L	$= Q/k_{32}$
CL_{24} = clearance of itraconazole by metabolism to hydroxyitraconazole	1.07 L/h	$= k_{24}V_2$
F_{12} = fraction of parent in the gut absolute bioavailability	0.55	(A)
F_{14} = fraction of metabolite after first pass metabolism	4.3×10^{-2}	(D)
F_{24} = fraction of parent converted to metabolite	0.5	(E)
F_{20} = fraction of parent eliminated	0.5	$= 1 - F_{24}$

differential equations. POPT® employs a version of simulated annealing for continuous variables similar to the one described in Section 2.4.1.

Numerical solutions to differential equations

In the case of the nonlinear model, there does not exist a closed form solution to Equations (4.5–4.8), so the responses of interest, A_2 and A_4 cannot be isolated analytically. Instead numerical solutions to the equations are used. For this purpose, MATLAB's built-in ODE solver `ode45` was employed. A relative tolerance (`RelTol`) of 10^{-4} was used in these calculations.

Numerical derivatives

Once the solutions to Equations (4.5–4.8) are found, their partial derivatives with respect to each parameter are required (as in Equation (4.26)). These derivatives were approximated numerically using the Central Difference Method,

$$\frac{\partial f(x)}{\partial x} \approx \frac{f(x(1+h_1)+h_2) - f(x(1-h_1)-h_2)}{2(xh_1+h_2)},$$

for a suitably small h_1 and h_2.

Since numerical approximations to both the ODE solutions and the partial derivatives are used, it is important that the ODE solutions are more precise than the numerical derivatives. For example, if the ODE solutions are correct to the 4^{th} decimal place, h_1 and h_2 must be chosen such that the changes in $f(\cdot)$ must be larger than 10^{-4}, otherwise the results may be inaccurate. An alternative solution would be to use the 'direct method' (see Atkinson and Bogacka (2002) for an example of its use).

4.4 Results

4.4.1 Optimal design

The set of optimal sampling times Ξ^* are given in Table 4.2. The times have been rounded to the nearest minute. As the sampling times have been chosen from a continuous interval,

these schedules are difficult to implement in practice. Indeed, in some cases they are actually impossible, with several blood samples being taken concurrently. It is therefore necessary to extend the population design to a sequence of intervals, or windows, in which the samples may be taken. We chose sampling windows of 3 hours (smaller for the earlier sampling times), roughly centered around the optimal times. These windows are also shown in Table 4.2.

The efficiency of a population design Ξ in this case is given by

$$\frac{D_P(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \Xi)}{D_P(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \Xi^*)}, \tag{4.29}$$

where Ξ^* is the optimal population design. Any deviation from the optimal sampling times will obviously lead to a decrease in efficiency. The sampling windows were evaluated by taking 1000 designs at random from all the given intervals and calculating their efficiencies. The 1000 designs were selected by taking a set of sampling times from uniform distributions on all of the sampling windows, i.e. these are joint (not marginal) sampling windows. Figure 4.3 shows the distribution of these efficiencies. The average efficiency was about 0.95, with 96.2% of the designs having an efficiency of 0.90 or greater. The 'worst' design found had an efficiency of about 0.85. The sampling windows were deemed acceptable in light of these results, and did not need to be shortened to give any further increase in efficiency.

4.4.2 Design evaluation

Model discrimination

The optimal design was evaluated in terms of its ability to select whether the 'true' underlying model involves linear or nonlinear elimination from the parent compartment.

To do this, 100 sets of data were simulated under each model (linear elimination and nonlinear elimination) using the parameter values given in Table 4.1 and the optimal design provided in Table 4.2. The models were fitted to each set of data using NONMEM (version 5 with the G77 FORTRAN compiler, using `FOCE` & `INTERACTION`), and the 'best' model was chosen empirically as the one with the smallest objective function (which is proportional to minus twice the log-likelihood). Recall that CL_{24}, Q and V_3 are considered to be fixed

Table 4.2: Optimal population design Ξ^* for the population pharmacokinetics study of itraconazole, with 10 patients in each of the 3 elementary design groups.

Group (q)	N_q	Capsule		Solution	
		Elementary design $\boldsymbol{\xi}_q^c$ (hrs:mins)	Sampling window (hrs)	Elementary design $\boldsymbol{\xi}_q^s$ (hrs:mins)	Sampling window (hrs)
1	10	1:14	0.1 → 3.0	0:17	0.1 → 1.0
		8:56	7.0 → 10.0	3:55	3.0 → 3.5
		25:49	24.0 → 27.0	3:56	3.5 → 4.0
		51:45	50.0 → 53.0	3:56	4.0 → 4.5
2	10	6:13	5.0 → 8.0	0:18	0.1 → 1.0
		9:50	8.0 → 11.0	4:06	3.0 → 4.0
		29:29	28.0 → 29.5	4:06	4.0 → 5.0
		29:29	29.5 → 31.0	72:00	69.0 → 72.0
3	10	8:08	7.0 → 10.0	0:17	0.1 → 1.0
		28:00	26.5 → 29.5	4:22	3.0 → 6.0
		72:00	69.0 → 70.5	27:08	26.0 → 29.0
		72:00	70.5 → 72.0	72:00	69.0 → 72.0

under both the linear and nonlinear models, and that K_m is considered fixed in the nonlinear elimination model.

After simulating under the linear model, the linear model was chosen correctly 74% of the time. After simulating under the nonlinear model, the nonlinear model was chosen correctly 100% of the time. For all estimation runs, the NONMEM run was reported to have converged successfully.

These results are considered to be acceptable, especially considering that at low concentrations (in comparison to K_m), the linear and Michaelis-Menten (nonlinear) models are virtually indistinguishable, meaning that the nonlinear elimination model will often fit very well to data generated by the linear model.

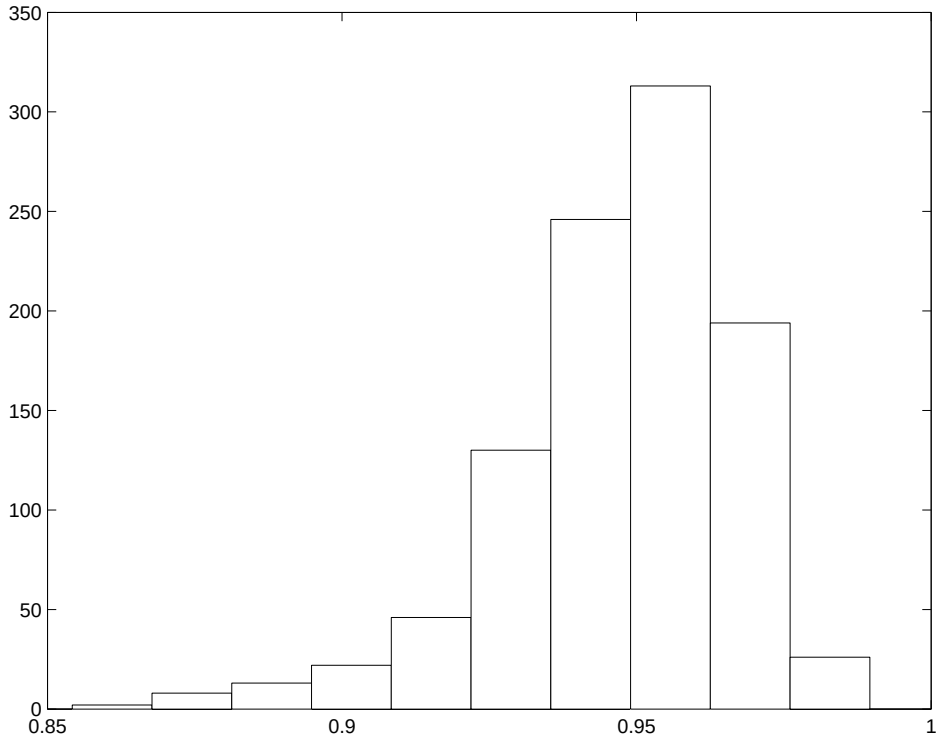

FIGURE 4.3: Histogram of efficiencies of 1000 simulations of sampling times taken uniformly within all of the sampling windows described in Table 4.2.

Standard errors

The results from these simulations were also used to calculate standard errors of the parameters estimates. Empirical estimates of the standard errors were obtained from the standard deviation of the estimated parameter values from the 100 simulated data sets. This was performed only in those cases where the same model was used for simulation and estimation. These are given in Table 4.3 for both the linear and nonlinear models. We can compare these standard errors to those predicted by the information matrix used to calculate the optimal design. These 'expected' standard errors are given by the square roots of the diagonals of the inverse of each Fisher information matrix. These are given in Table 4.3.

The comparison of expected and simulated standard errors reveal quite satisfactory results, considering the difficulty in estimating between subject variability and absorption rate constants. The only parameter with a significant difference between the expected and

Table 4.3: Standard errors (expressed as coefficients of variation) for the population mean and between subject variability for parameters estimated under the linear and nonlinear models. The expected standard errors are calculated from the Fisher information matrix, the estimated standard errors are the standard deviations of parameter estimates from 100 NONMEM simulation-estimation runs. The dash indicates that normality was rejected by the Kolmogorov-Smirnov test (the data was highly positively skewed), in which case calculation of standard errors in meaningless.

	Model parameter				
Linear elimination of parent	k_a	V_2	V_4	CL_4	CL_2
Population mean					
Expected CV (%)	8.39	16.39	10.64	10.65	8.57
Estimated CV (%)	54.34	18.71	18.20	10.76	8.32
Between subject variability					
Expected CV (%)	41.22	82.17	51.57	45.58	39.95
Estimated CV (%)	121.00	114.61	80.85	68.30	89.80
Nonlinear elimination of parent	k_a	V_2	V_4	CL_4	$V_{\max}$
Population mean					
Expected CV (%)	8.19	15.55	11.26	12.51	27.84
Estimated CV (%)	25.25	10.30	9.85	10.58	33.44
Between subject variability					
Expected CV (%)	40.81	49.04	49.28	61.62	197.73
Estimated CV (%)	111.00	76.10	83.00	81.10	—

estimated standard errors is k_a. The absorption rate constant is notoriously difficult to estimate in PK models, so it is not unexpected that the estimates from the simulations were so variable.

Parameter estimation

To assess the benefit of the optimal population design with respect to parameter estimation, two additional designs were calculated: one is optimal in terms of the linear model alone, and the other optimal in terms of the nonlinear model alone. The D-optimal design for model u, denoted by Ξ_u^*, was found by maximising the criterion

$$|\boldsymbol{M}^u(\boldsymbol{\Psi}^u, \Xi)|^{1/p_u}, \quad u = 1, 2 \tag{4.30}$$

using the same methods as described in Section 4.3.5.

The efficiency of any population design Ξ with respect to model u is defined (in a similar fashion to Equation (4.29)) to be the ratio

$$\text{Eff}_u(\Xi, \Xi_u^*) = \frac{|\boldsymbol{M}^u(\boldsymbol{\Psi}^u, \Xi)|^{1/p_u}}{|\boldsymbol{M}^u(\boldsymbol{\Psi}^u, \Xi_u^*)|^{1/p_u}} \quad u = 1, 2.$$

where Ξ^* is the product optimal design given in Table 4.2.

For the product optimal design Ξ^*, the criteria given by Equation (4.30) were calculated under the linear and nonlinear models separately. The values found were 37.591 and 115.1, respectively. The criteria for designs Ξ_1^* and Ξ_2^*, which are optimal under the linear and nonlinear model, respectively, are 38.993 and 120.01. Hence the efficiencies of Ξ^* in terms of the linear and nonlinear models are:

$$\text{Eff}_1(\Xi^*, \Xi_1^*) = \frac{37.591}{38.993} = 0.9640$$

and

$$\text{Eff}_2(\Xi^*, \Xi_2^*) = \frac{115.1}{120.01} = 0.9591.$$

This indicates that the product optimal design is suitably efficient at estimating the parameters of both models.

4.5 Discussion

The practical example of the itraconazole study is still underway at the time of writing. The authors are currently waiting on results from the actual trial.

For comparison, the empirical 'rich' sampling design given in Section 4.3.3 was used, and the expected standard errors from the design were calculated from the information matrix. In this case, again in a study involving thirty patients, all parameters in question would be estimable. The relative standard error of all fixed effects would all be less than 20%, in most cases less than 10%.

The construction of the optimal design was subject to several constraints. A timeline of approximately two months was given to develop the entire design, from collating prior information on itraconazole and researching applicable methods, to computing the final design (a procedure which took over seven days in itself). Future research in this area includes

the consideration of alternative methodology (such as the 'direct method' of Atkinson and Bogacka (2002) and more sophisticated incorporation of multiple response types for mixed effects models). Under the given time constraints, computation time was a major issue, with the final design computation taking over a week to run (on a 2.4 GHz Pentium 4 with 1 GB of RAM). Cost limitations of the study and limitations imposed by the relevant ethics committee also limited the possible number of patients and the number of samples per patient, which further constrained the design in terms of the number of parameters which are estimable.

Several ideas were used in this design which are novel to the field of experimental design for population pharmacokinetic studies. These include the use of the D-optimality product criterion for model discrimination and the incorporation of multiple responses into the Fisher information matrix.

The results of the power tests and simulations show that the use of this product criterion has created a design which is efficient in terms of both parameter estimation and model discrimination. Whether the individual D-optimal designs were also adequate for model discrimination was not investigated for this design. However, we have seen evidence (Waterhouse *et al.*, 2004b) that suggests that this is not true for nonlinear models.

The standard errors predicted by the inverse of the Fisher information matrix generally match those estimated by the simulations. We believe that these results show that the approximations made by Retout and Mentré (2003) are mostly inconsequential. Also, the assumptions about the covariance structure of the two responses in the model seem to be justified by these results.

The sparse sampling design presented in this paper gives a simple and cost-effective data collection regimen which should provide acceptable estimates of the parameters of interest.

Part II

Generalised linear models

The ideas formed in Part I of this thesis are now carried over to generalised linear models. GLMs present some additional challenges in regard to discrimination of nested models. The problem of constructing designs for both effective discrimination between two competing nested models and efficient estimates of the parameters of both models is addressed, and demonstrated for a number of examples. The optimal design of crossover studies in another pharmacological application is also investigated, where the structural model is assumed to be known. The sensitivity to parameter misspecification of the constructed designs is then examined. Finally, we extend the logistic regression model used in the crossover study to include random coefficients and evaluate some ideas for optimal design for these models, where design for estimation remains a challenge.

Chapter 5

Designs for discrimination and estimation in generalised linear models

5.1 Introduction

The optimal design of experiments for efficient parameter estimation in generalized linear models (GLMs), as for the nonlinear models discussed in Chapter 3, may be addressed through the use of well-established criteria such as the D-optimality criterion (see Atkinson and Haines (1996) for an overview of D-optimal designs for GLMs), although results in this area are generally limited to models with no more than two explanatory variables. Model discrimination for GLMs, on the other hand, is a problem in optimal design which has received relatively little attention. GLMs present additional challenges to the problem of model discrimination, which are addressed in this chapter.

The T-optimality criterion for optimal design for discrimination between two nonlinear models was defined in Atkinson and Fedorov (1975a), and examples of its use have been shown in Chapter 3 of this thesis. Ponce de Leon and Atkinson (1992) subsequently extended this criterion to apply to generalised linear models of the same subclass, i.e. models where the responses share the same distribution class. The criterion involves maximising the deviance

(as defined in Section 2.2.3) arising from the fit of one model, where the data are generated under another model which is assumed to be true. Müller and Ponce de Leon (1996) describe a sequential approach to the design problem, again using the maximum deviance of the 'untrue' model to choose each additional design point.

However, the deviance function may not always be a suitable measure of goodness-of-fit for binary data (McCullagh and Nelder, 1989). Under certain conditions, the asymptotic χ^2_{n-p} distribution (where n is the number of support points, and p is the number of parameters in the model under consideration) does not hold, and the deviance is not independent of the fitted probabilities, in which case a large deviance "cannot necessarily be considered to be evidence of a poor fit" (McCullagh and Nelder, 1989, p.119). McCullagh and Nelder point out, though, that the deviance function can be useful for comparing two nested models. The χ^2 distribution is "usually quite accurate" for differences in deviances of two nested models (McCullagh and Nelder, 1989, p.119), and we use this as a basis for criteria for designs to discriminate between such models.

This chapter will assess the ability of T-optimal designs to discriminate between two nested logistic regression models, augment these designs with product optimal designs support points to improve parameter estimation (as described in Chapter 3), and suggest an alternative optimality criterion based on the expected difference in deviance of the models.

5.2 Methods

5.2.1 Criteria

T-optimality

The T-optimality criterion is now restated for GLMs with binary responses, as given in Section 2.3.3, in slightly more detail. Suppose that we have an n-point approximate design

$$\xi = \begin{Bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix},$$

and that at each support point of the design, we observe the number of successes S_i, which has a binomial distribution $\text{Bin}(m_i, \pi_i)$. The proportion of successes $Y_i = S_i/m_i$ may be

modelled by a GLM. Recall from Section 2.2 that a GLM for binomial data is given by

$$g(\pi_i) = \eta_i, \quad i = 1, \ldots, n,$$

and

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)' = \boldsymbol{X}\boldsymbol{\beta},$$

where each row of the $n \times p$ matrix $\boldsymbol{X}$ is a p-vector of known functions of $\boldsymbol{\xi}_i$; $\boldsymbol{\beta}$ is the $p \times 1$ vector of model parameters; and the link function g is usually either the logistic link function, $g(\pi_i) = \log\{\pi_i/(1-\pi_i)\}$; the probit link function, $g(\pi_i) = \Phi^{-1}(\pi_i)$, where Φ is the cumulative distribution function of the standard normal distribution; or the complementary log-log link function, $g(\pi_i) = \log\{-\log(1-\pi_i)\}$.

The log-likelihood function is

$$\ell(\boldsymbol{\pi};\, \boldsymbol{Y}) = \sum_{i=1}^{n} w_i \left[Y_i \log \pi_i + (1 - Y_i) \log(1 - \pi_i) \right],$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)'$.

Suppose that we are interested in choosing between two candidate GLMs, M_1 and M_2 with the same link function but different linear predictors, $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ respectively, where $\boldsymbol{\eta}_1$ is 'nested' in $\boldsymbol{\eta}_2$, i.e. $\boldsymbol{\eta}_2$ contains all of the terms from $\boldsymbol{\eta}_1$ as well as one or more extra covariates. Let the parameter vectors for models M_1 and M_2 be the p_1-vector $\boldsymbol{\beta}_1$ and the p_2-vector $\boldsymbol{\beta}_2$, respectively. Denote the maximum likelihood estimates of the probabilities from each model as $\hat{\boldsymbol{\pi}}_1 = (\hat{\pi}_{11}, \ldots, \hat{\pi}_{1n})'$ and $\hat{\boldsymbol{\pi}}_2 = (\hat{\pi}_{21}, \ldots, \hat{\pi}_{2n})'$. The deviance from the fit of model u (with $u = 1$ or 2) is defined as

$$\begin{aligned} D(\boldsymbol{Y}; \hat{\boldsymbol{\pi}}_u) &= 2\ell(\boldsymbol{Y};\, \boldsymbol{Y}) - 2\ell(\hat{\boldsymbol{\pi}}_u;\, \boldsymbol{Y}) \\ &= 2\sum_{i=1}^{n} w_i \left[Y_i \log(Y_i/\hat{\pi}_{ui}) + (1 - Y_i) \log\left(\frac{1 - Y_i}{1 - \hat{\pi}_{ui}} \right) \right]. \end{aligned}$$

For a T-optimal design, the assumption is made that the larger model M_2 is the 'correct' model with $\boldsymbol{\pi}_2$ known. The T-optimality criterion is the deviance arising from the fit of M_1 to the expected value of $\boldsymbol{Y}$ under M_2, $\boldsymbol{\pi}_2$. That is, the criterion is

$$D(\boldsymbol{\pi}_2; \hat{\boldsymbol{\pi}}_1). \tag{5.1}$$

Define the T-efficiency of a design as the ratio of its value of T-optimality criterion, as given above in Equation (5.1), to the T-optimality criterion value for the T-optimal design. This is analogous to the T^b_{eff} efficiency from Equation (3.3).

T_E-optimality

As mentioned previously, the random variable $D(\boldsymbol{Y}; \hat{\boldsymbol{\pi}}_u)$ may not always follow the asymptotic $\chi^2_{n-p_u}$ distribution, and so may be unsuitable as a measure of goodness of fit, and hence unsuitable as a design criterion. The reduction in deviance may be considered instead:

$$\begin{aligned} R(\boldsymbol{Y}) &= D(\boldsymbol{Y}; \hat{\boldsymbol{\pi}}_1) - D(\boldsymbol{Y}; \hat{\boldsymbol{\pi}}_2) \\ &= 2\ell(\hat{\boldsymbol{\pi}}_2;\, \boldsymbol{Y}) - 2\ell(\hat{\boldsymbol{\pi}}_1;\, \boldsymbol{Y}). \end{aligned} \tag{5.2}$$

Under M_1, $R(\boldsymbol{Y})$ is distributed approximately as $\chi^2_{p_2-p_1}$ under the assumptions of no overdispersion and fixed n.

As an alternative to T-optimal designs for GLMs, consider the *average* reduction in deviance given that the second model is true, i.e. an alternative criterion is $\mathrm{E}(R(\boldsymbol{Y})|M_2)$. As this expectation may not be evaluated analytically, Monte Carlo simulation is used to produce an approximation to the criterion:

$$E(R(\boldsymbol{Y})|M_2) \approx \sum_{i=1}^{N_s} R(\boldsymbol{Y}_i^*)/N_s, \tag{5.3}$$

where N_s is the number of simulations (typically large, at least 10^3) and $\boldsymbol{Y}_i^* = (Y_{i1}^*, \ldots, Y_{in}^*)^T$ are the simulated data. The simulated data in this case are the sample proportions, taken from a truncated normal distribution with mean $\boldsymbol{\pi}_2$ and a variance based on the structure of the binomial variance,

$$\mathrm{Var}(Y_{ij}^*) = \pi_{2j}(1 - \pi_{2j})K,$$

where K is chosen to quantify the experimenter's confidence in M_2 (a larger K would result in increased variation from M_2, so would correspond to less confidence in the larger model). We refer to a design which maximises this optimality criterion in Equation (5.3) as T_E-optimal.

The T_E-optimality criterion has desirable theoretical qualities in terms of constructing designs for model discrimination, as the χ^2 distribution of the reduction in deviance is

expected to be more accurate than that of the single deviance used in T-optimality. However, due to the large number of simulations required to produce the criterion value for a single design, it is not feasible to be used to construct an optimal design, especially when using a search algorithm like simulated annealing, which typically requires hundreds of thousands of calculations of the criterion. Instead focus is placed on evaluation of the T-optimal designs described above and the construction of hybrid designs described below.

Hybrid designs

A compromise between model discrimination and parameter estimation is sought, as in Chapter 3. To achieve this objective, the hybrid designs described in that chapter may be used, as they produce similar results to the conditional designs with much less computational effort. This involves finding the T-optimal design, as well as the corresponding product optimal design, and combining the design points of the two designs, with weights reflecting the importance placed on each objective.

5.2.2 Power tests

The ability of an experimental design to discriminate between two competing models may be assessed by power tests. In these power tests, a large number (N_{sim}) of sets of simulated random binomial data $\boldsymbol{s}_k^*$ (a vector of size n for an n-point design, for $k = 1, \ldots, N_{\text{sim}}$) is generated using the true model M_2. For a given design ξ, each simulated vector $\boldsymbol{s}_k^*(\xi) = (s_{k1}^*(\xi), \ldots, s_{kn}^*(\xi))'$ is a realisation of $\boldsymbol{S}^*(\xi) = (S_1^*(\xi), \ldots, S_n^*(\xi))'$, whose elements have the following distribution:

$$S_i^*(\xi) \sim \text{Bin}([w_i N_{\text{sample}}], \pi_{2i}),$$

where $[x]$ is the integer nearest to x. N_{sample} is the total size of the sample from which the binomial counts are taken in the simulations, chosen in an ad hoc manner to be large enough to be able to reasonably fit the models to the data, but not so large that discriminating between the two models becomes a trivial exercise. The simulated proportions of success $\boldsymbol{y}_k^*(\xi) = (y_{k1}^*(\xi), \ldots, y_{kn}^*(\xi))'$ are then calculated by $y_{ki}^*(\xi) = s_{ki}^*(\xi)/[w_i N_{\text{sample}}]$.

The power of the design is then defined to be the proportion of times that model M_2 is

correctly chosen as a better fit to the data than model M_1. The choice of the 'best' fit is determined by carrying out the likelihood ratio test for testing $H_0 : M_1$ against $H_A : M_2$. The test statistic is the reduction in deviance $R(\boldsymbol{y}_k^*(\xi))$ as defined in Equation (5.2), which is distributed approximately like $\chi^2_{p_2-p_1}$. We select M_2 as the more appropriate model on the kth simulation if $R(\boldsymbol{y}_k^*(\xi)) > \chi^2_{p_2-p_1;0.95}$.

5.3 Nested logistic regression models with two factors

A likely scenario for model discrimination with GLMs is the choice between logistic regression models with two factors, where interaction between the two factors is included in one model and omitted in the other (so $p_1 = 3$ and $p_2 = 4$).

$$\begin{aligned} M_1 : \quad & \text{logit}(\pi_1) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 \\ M_2 : \quad & \text{logit}(\pi_2) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_{i1}x_2 \\ & -1 \leq x_1, x_2 \leq 1 \end{aligned}$$

The choice of bounds for x_1 and x_2 was made to 'standardise' the design region. Any bounded explanatory variable may be scaled to the same region, as long as the model coefficients are altered accordingly.

5.3.1 Parameter values

As all of the design criteria presented in this chapter rely on the specification of the parameters for the second model, a value must be set for $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})'$. β_{20}, β_{21} and β_{22} are all arbitrarily set to 1. (The next section considers the case where β_{22}, the coefficient of x_2, is twice that of the coefficient of x_1, i.e. $\beta_{22} = 2$. The first set of results will hence be labelled as the '$\beta_{22} = 1$ case'.) A range of values for β_{23} (the interaction coefficient) is chosen so that the interaction creates a big enough difference between the two models to make it worthwhile designing an experiment to discriminate between them, but not so large as to make the discrimination problem trivial.

For a range of values of the interaction coefficient (from -4 to 4), the expected responses under model M_2 ($\boldsymbol{\pi}_2$) was calculated on a 20×20 grid of values of x_1 and x_2, each over

the range $[-1, 1]$. The smaller model M_1 was then fitted to these expected responses using iteratively reweighted least squares. Denote the fitted probabilities by $\hat{\boldsymbol{\pi}}_1$ and the fitted parameter vectors by $\hat{\boldsymbol{\beta}}_1$. The differences $\boldsymbol{\pi}_2 - \hat{\boldsymbol{\pi}}_1$ are shown in Figure 5.1, along with the values of $\boldsymbol{\beta}_2$ and $\hat{\boldsymbol{\beta}}_1$ in each case. After examining these plots, it was decided that the most interesting range of values of the interaction coefficient was from -1 to 1. Outside this range, the difference between the models created by the large interaction would make model discrimination trivial.

5.3.2 Designs for discrimination

The T-optimal designs for the various interaction parameter values are given in Table 5.1 and Figure 5.2. Each support point in the figure is represented by its position in the xy-plane, and its corresponding weight is proportional to the size (area) of the plotted point.

The actual support points for the T-optimal designs do not change for different values of the interaction coefficient, they are always at the corners of the design region, i.e. the support points of the 2^2 factorial design. Only the weights are affected by the value of this parameter, and these changes are only minimal. The differences in weights between these designs and the factorial design (with equal weights of 0.25), however, is important. The 2^2 factorial design has reasonably lower T-efficiencies than these T-optimal designs, ranging from 0.88 to 0.93.

The fitted parameter values for model M_1 under the T-optimal designs are given in Table 5.2.

5.3.3 Designs for parameter estimation

Assessment of the T-optimal designs above in terms of their ability to produce good parameter estimates was also considered. To do this, locally D-optimal designs were generated to compare the T-optimal designs against, in which case values of $\boldsymbol{\beta}_1$ need to be specified ($\boldsymbol{\beta}_2$ is already known from the construction of the T-optimal designs). The MLE estimates of $\boldsymbol{\beta}_1$ from the T-optimal designs found previously, given in Table 5.2, were used here.

D-optimal designs for models M_1 and M_2 are given in Tables 5.3 and 5.4, as well as

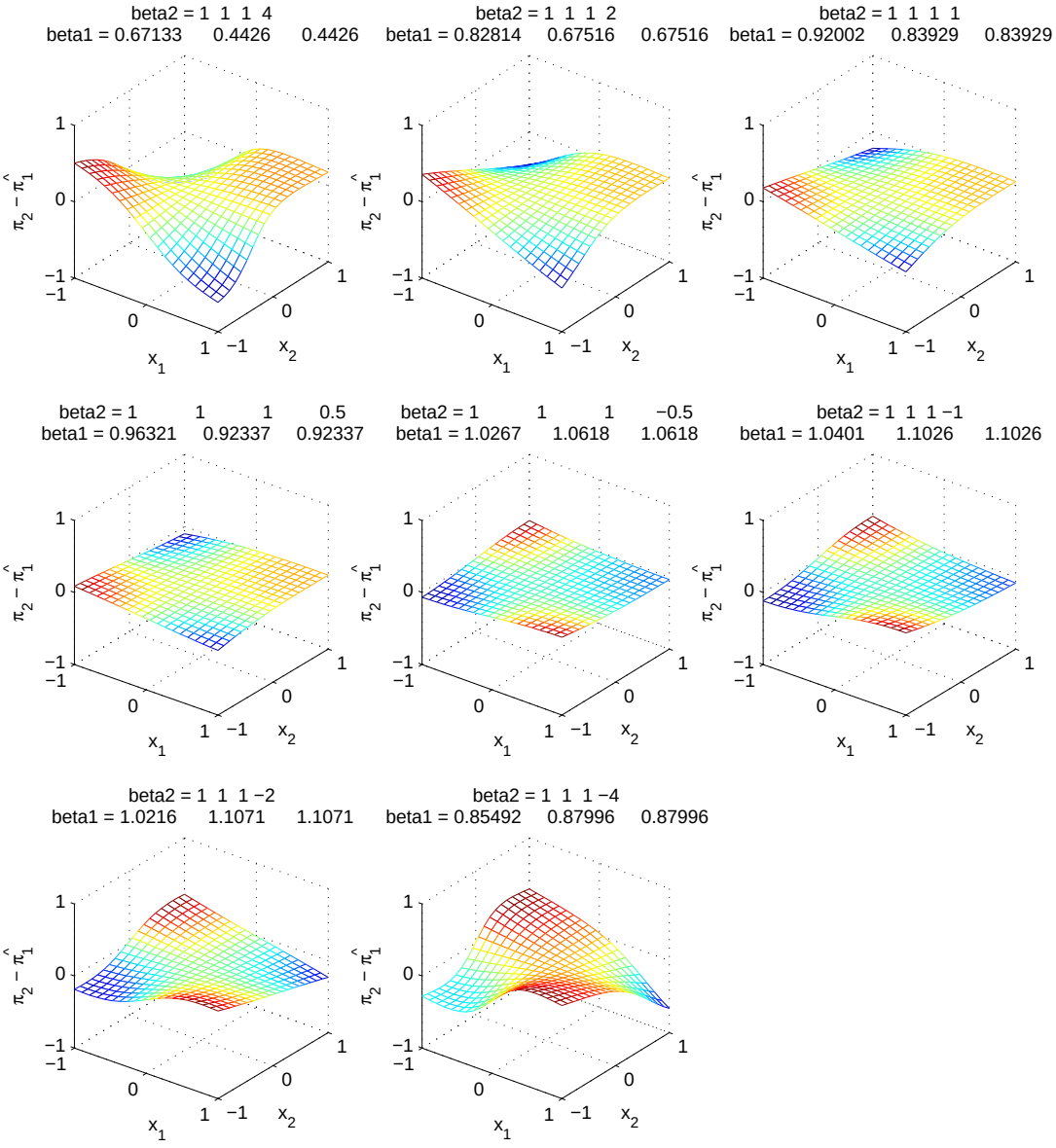

FIGURE 5.1: The difference between models M_1 and M_2 over various values of the interaction coefficient in M_2 when $\beta_{22} = 1$. For a given $\boldsymbol{\beta}_2$, $\hat{\boldsymbol{\beta}}_1$ is the best fit of M_1 to the expected responses under M_2. The colour of the surface is indicative of the difference between the expected response of the models: red signifies that π_2 is greater than $\hat{\pi}_1$, blue signifies that π_2 is less than $\hat{\pi}_1$.

Table 5.1: T-optimal designs for nested logistic models with two variables when $\beta_{22} = 1$.

$\boldsymbol{\beta}_2'$		T-optimal design			
$(1, 1, 1, 1)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	1.0000
	w	0.2038	0.2038	0.2038	0.3886
$(1, 1, 1, 0.5)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	1.0000
	w	0.1973	0.1973	0.1973	0.4081
$(1, 1, 1, 0.3)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	1.0000
	w	0.1962	0.1962	0.1962	0.4113
$(1, 1, 1, -0.3)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	1.0000
	w	0.1994	0.1994	0.1994	0.4017
$(1, 1, 1, -0.5)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	1.0000
	w	0.2025	0.2025	0.2025	0.3924
$(1, 1, 1, -1)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	1.0000
	w	0.2132	0.2132	0.2132	0.3604

Table 5.2: Fitted values of the parameters of the smaller model, given by the T-optimal designs for the various values of $\boldsymbol{\beta}_2$, when $\beta_{22} = 1$.

$\boldsymbol{\beta}_2'$	$\hat{\boldsymbol{\beta}}_1'$
$(1, 1, 1, 1)$	$(0.7072, 0.7072, 0.7072)$
$(1, 1, 1, 0.5)$	$(0.8679, 0.8679, 0.8679)$
$(1, 1, 1, 0.3)$	$(0.9261, 0.9261, 0.9261)$
$(1, 1, 1, -0.3)$	$(1.0523, 1.0523, 1.0523)$
$(1, 1, 1, -0.5)$	$(1.0728, 1.0728, 1.0728)$
$(1, 1, 1, -1)$	$(1.0707, 1.0706, 1.0706)$

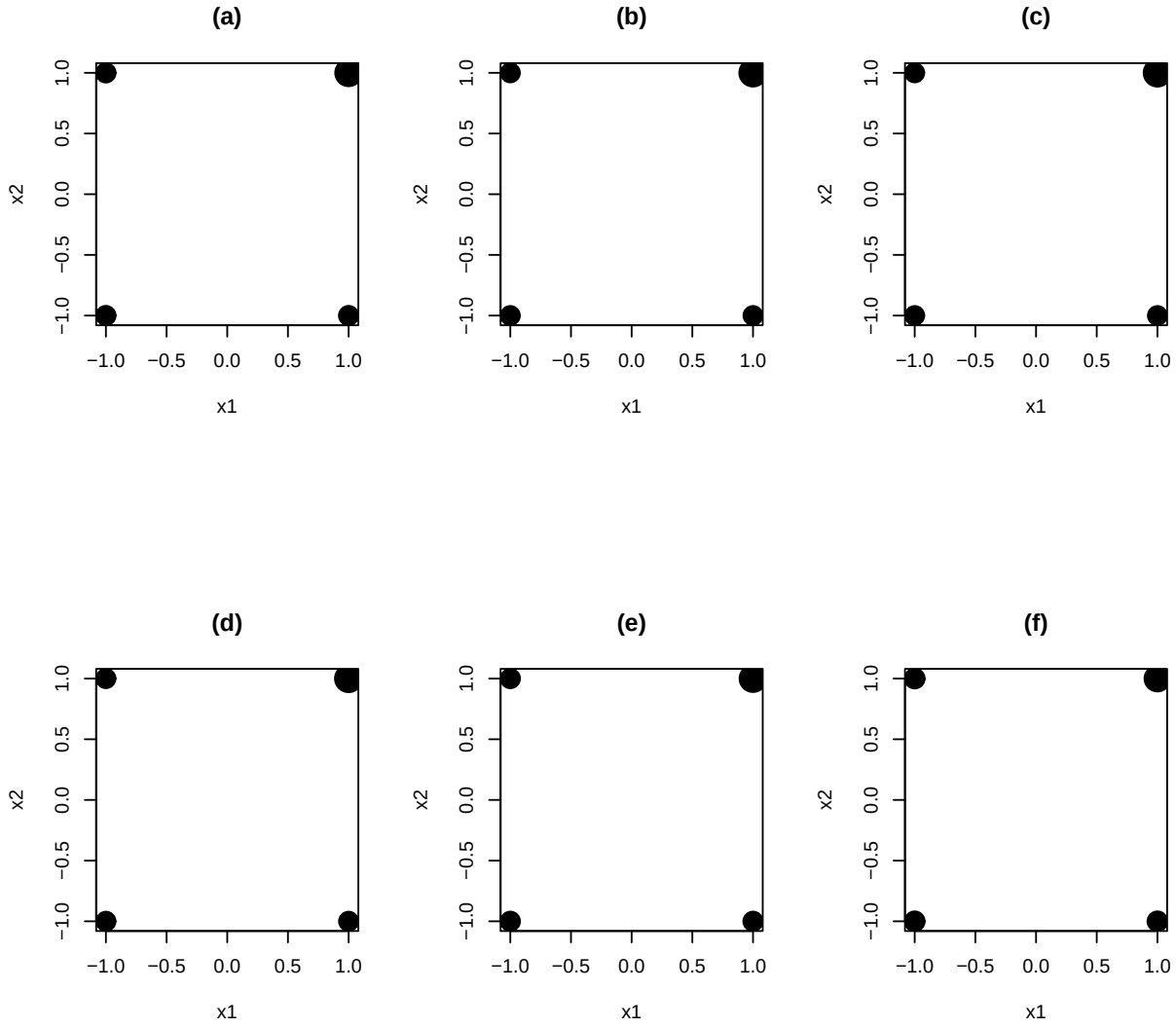

Figure 5.2: T-optimal designs for (a) $\boldsymbol{\beta}_2 = (1, 1, 1, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 1, -1)'$.

in Figures 5.3 and 5.4. The product of the D-optimality criteria was also used to find the 'product optimal' designs (as described in Section 3.1.2 for nonlinear models), which are given in Table 5.5 and Figure 5.5.

The first D-optimal design for model M_1 presented here (when $\beta_{23} = 1$) shares the same support points as the T-optimal design, with different weights. For lower values of this parameter, however, the point (1,1) disappears to leave a three-point design at the remaining corners of the design space.

More interesting results are seen for the D-optimal designs for model M_2, where for β_{23} equal to -0.3 or -0.5 the support points are again in the corners of the design region, but for different values of β_{23} there is either a further deviation from the point (1,1), or an increase in the number of support points to 5.

The product optimal designs show a compromise between the D-optimal designs for models M_1 and M_2. The three corners $(-1,-1)$, $(-1,1)$ and $(1,-1)$ (or points very nearby) are again common characteristics of the design, with the position of the remaining point(s) seemingly dictated by the corresponding D-optimal designs for model M_2. Comparing these to the T-optimal designs seems to suggest that for greater positive values of the interaction coefficient, moving design points away from (1,1) seems to produce designs more efficient in regard to parameter estimation.

5.3.4 Hybrid designs and power tests

Given the T-optimal designs in Section 5.3.2 and the product optimal designs in Section 5.3.3, these are combined to form hybrid designs, using the procedure outlined in Section 3.1.4. Hybrid designs are calculated for many levels of α, from $\alpha = 0$ (T-optimal design) up to $\alpha = 1$ (product optimal design), in steps of 0.05. For each design, its T-efficiency, its D-efficiencies under both models, and its power to discriminate between the two models are calculated, using the method described in Section 5.2.2.

For the power tests, the value of N_{sample} was chosen by some preliminary investigation, where power tests were performed on a number of hybrid designs for all parameter values. $N_{\text{sample}} = 200$ seemed to be large enough that both models would fit the simulated data at

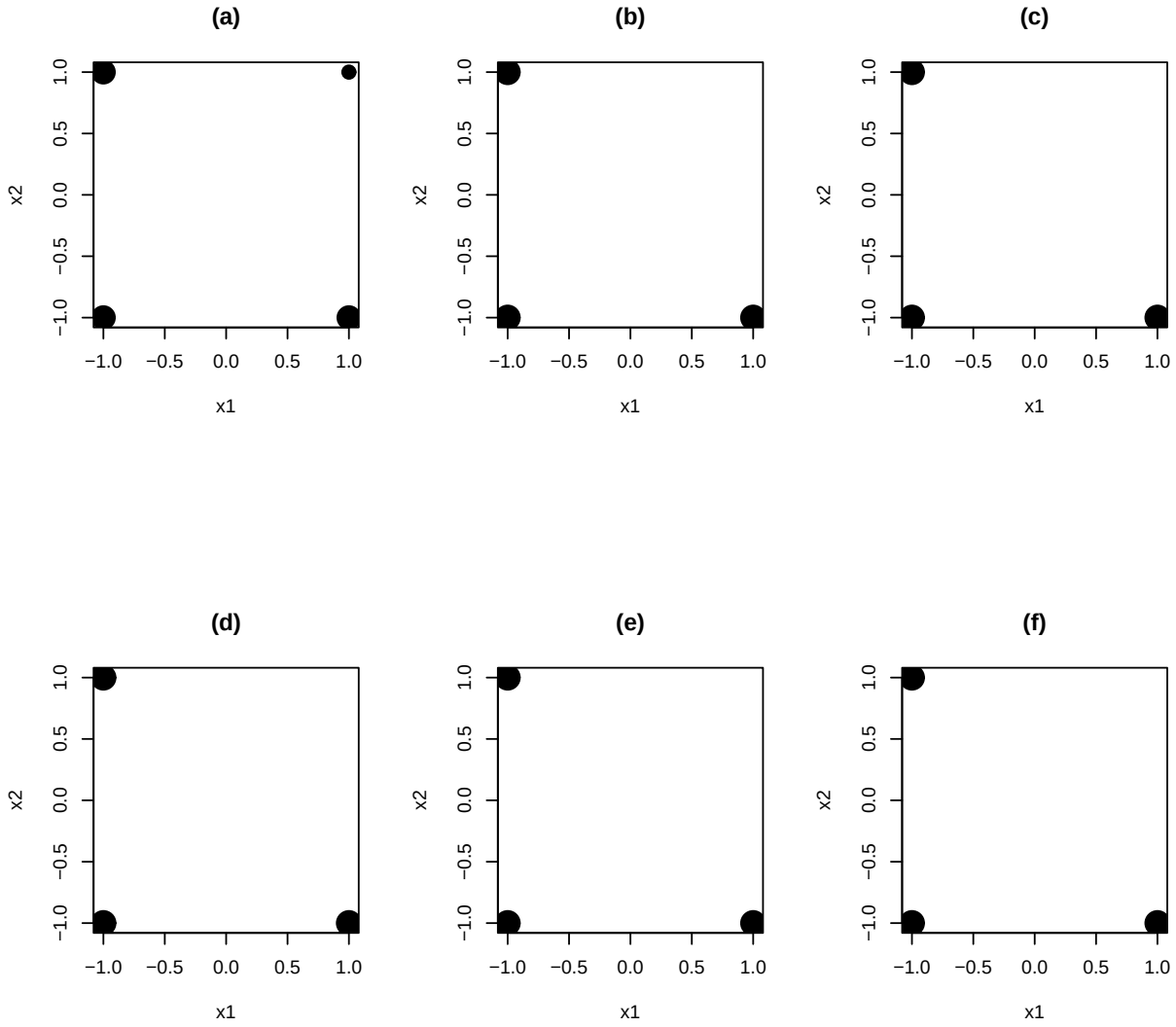

Figure 5.3: D-optimal designs for model M_1, for (a) $\boldsymbol{\beta}_1 = (0.7072, 0.7072, 0.7072)'$, (b) $\boldsymbol{\beta}_1 = (0.8679, 0.8679, 0.8679)'$, (c) $\boldsymbol{\beta}_1 = (0.9261, 0.9261, 0.9261)'$, (d) $\boldsymbol{\beta}_1 = (1.0523, 1.0523, 1.0523)'$, (e) $\boldsymbol{\beta}_1 = (1.0728, 1.0728, 1.0728)'$, and (f) $\boldsymbol{\beta}_1 = (1.0707, 1.0706, 1.0706)'$. The values of $\hat{\boldsymbol{\beta}}_1$ are taken from the T-optimal designs for the values of $\boldsymbol{\beta}_2$ given in Table 5.2.

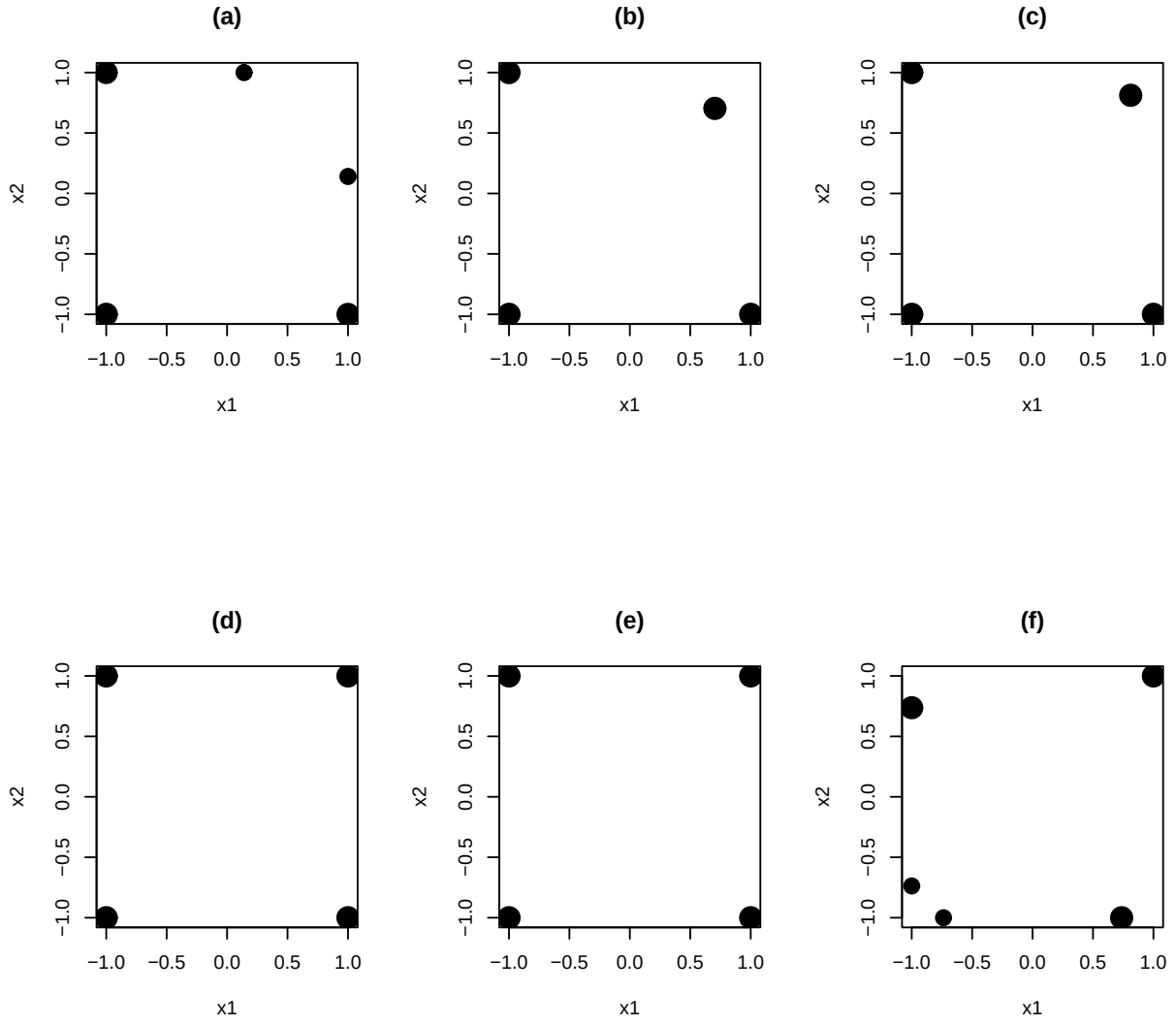

Figure 5.4: D-optimal designs for model M_2, for (a) $\boldsymbol{\beta}_2 = (1, 1, 1, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 1, -1)'$.

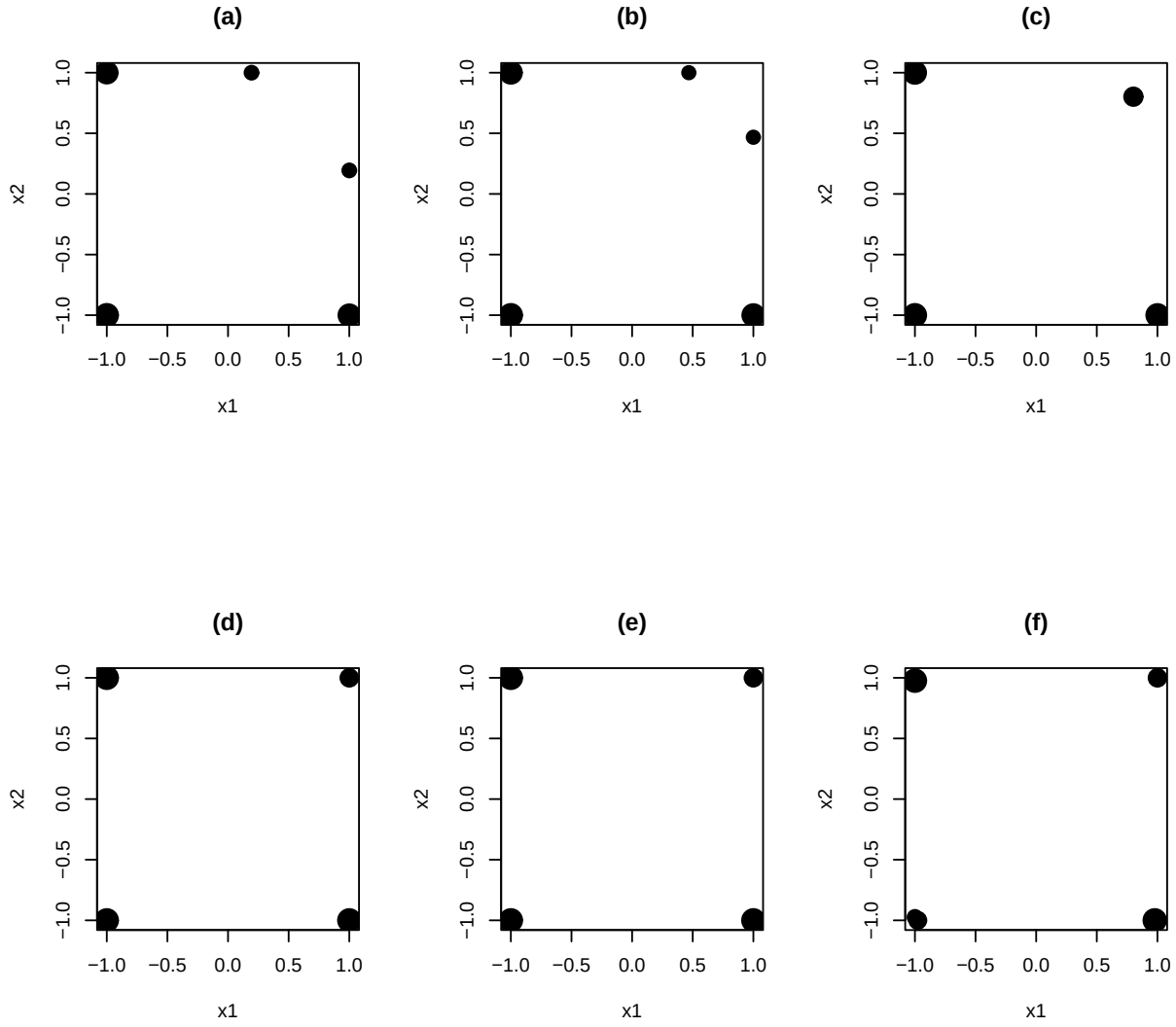

FIGURE 5.5: Product optimal designs for models M_1 and M_2, for (a) $\boldsymbol{\beta}_2 = (1, 1, 1, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 1, -1)'$. Corresponding values of $\hat{\boldsymbol{\beta}}_1$ for these values of $\boldsymbol{\beta}_2$ are given in Table 5.2.

Table 5.3: D-optimal designs for model M_1 when $\beta_{22} = 1$. The values of $\hat{\boldsymbol{\beta}}_1$ are taken from the T-optimal designs for the values of $\boldsymbol{\beta}_2$ given in Table 5.2.

$\hat{\boldsymbol{\beta}}_1'$		D-optimal design			
$(0.7072, 0.7072, 0.7072)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	1.0000
	w	0.2991	0.2991	0.2991	0.1026
$(0.8679, 0.8679, 0.8679)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-1.0000	1.0000	-1.0000	
	w	0.3333	0.3333	0.3333	
$(0.9261, 0.9261, 0.9261)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-1.0000	1.0000	-1.0000	
	w	0.3333	0.3333	0.3333	
$(1.0523, 1.0523, 1.0523)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-1.0000	1.0000	-1.0000	
	w	0.3333	0.3333	0.3333	
$(1.0728, 1.0728, 1.0728)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-1.0000	1.0000	-1.0000	
	w	0.3333	0.3333	0.3333	
$(1.0707, 1.0706, 1.0706)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-1.0000	1.0000	-1.0000	
	w	0.3333	0.3333	0.3333	

least about 97% of the time (data sets where both models did not fit were not included in the N_{sim} data sets used to calculate the power), but still not so large that the power was always close to 1, regardless of the design. $N_{\text{sim}} = 5000$ was chosen to give reasonably stable results. The results for the various parameter values are shown in Figure 5.6.

As seen in the examples in Chapter 3, it can be seen that as α increases (reflecting an increasing importance placed on parameter estimation), the D-efficiencies increase, and the T-efficiency decreases.

The results of the power tests are of more interest here. It is important to note that the size of the power is not of great concern here, as the power of a design is highly dependent on the size of the samples used in the simulations. Rather it is the trend in power as α

Table 5.4: D-optimal designs for model M_2 when $\beta_{22} = 1$.

$\boldsymbol{\beta}_2'$		D-optimal design				
$(1, 1, 1, 1)$	x_1	-1.0000	-1.0000	0.1404	1.0000	1.0000
	x_2	-1.0000	1.0000	1.0000	-1.0000	0.1403
	w	0.2500	0.2461	0.1289	0.2461	0.1290
$(1, 1, 1, 0.5)$	x_1	-1.0000	-1.0000	0.7036	1.0000	
	x_2	-1.0000	1.0000	0.7034	-1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 1, 0.3)$	x_1	-1.0000	-1.0000	0.8119	1.0000	
	x_2	-1.0000	1.0000	0.8118	-1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 1, -0.3)$	x_1	-1.0000	-1.0000	1.0000	1.0000	
	x_2	-1.0000	1.0000	-1.0000	1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 1, -0.5)$	x_1	-1.0000	-1.0000	1.0000	1.0000	
	x_2	-1.0000	1.0000	-1.0000	1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 1, -1)$	x_1	-1.0000	-1.0000	-0.7370	0.7370	1.0000
	x_2	-0.7370	0.7370	-1.0000	-1.0000	1.0000
	w	0.1264	0.2486	0.1264	0.2486	0.2500

changes that is of interest. The slightly irregular shape of the power plots is due to the random nature of the power tests. Increasing N_{sim} would smooth these curves somewhat.

When the magnitude of the interaction coefficient β_{23} is relatively large (1 or -1), the two models are very easy to distinguish between for these sample sizes, so the powers are nearly always very close to 1. The trends of the powers are easier to interpret for $|\beta_{23}| < 1$. It seems that an increase in α results in a general decrease in power. This shows that putting more weights on the T-optimal design points results in a design with greater power to discriminate between the models. This increase in power, however, is almost negligible. For these parameter values, the product designs show a power to discriminate between the models almost equivalent to that of the T-optimal designs.

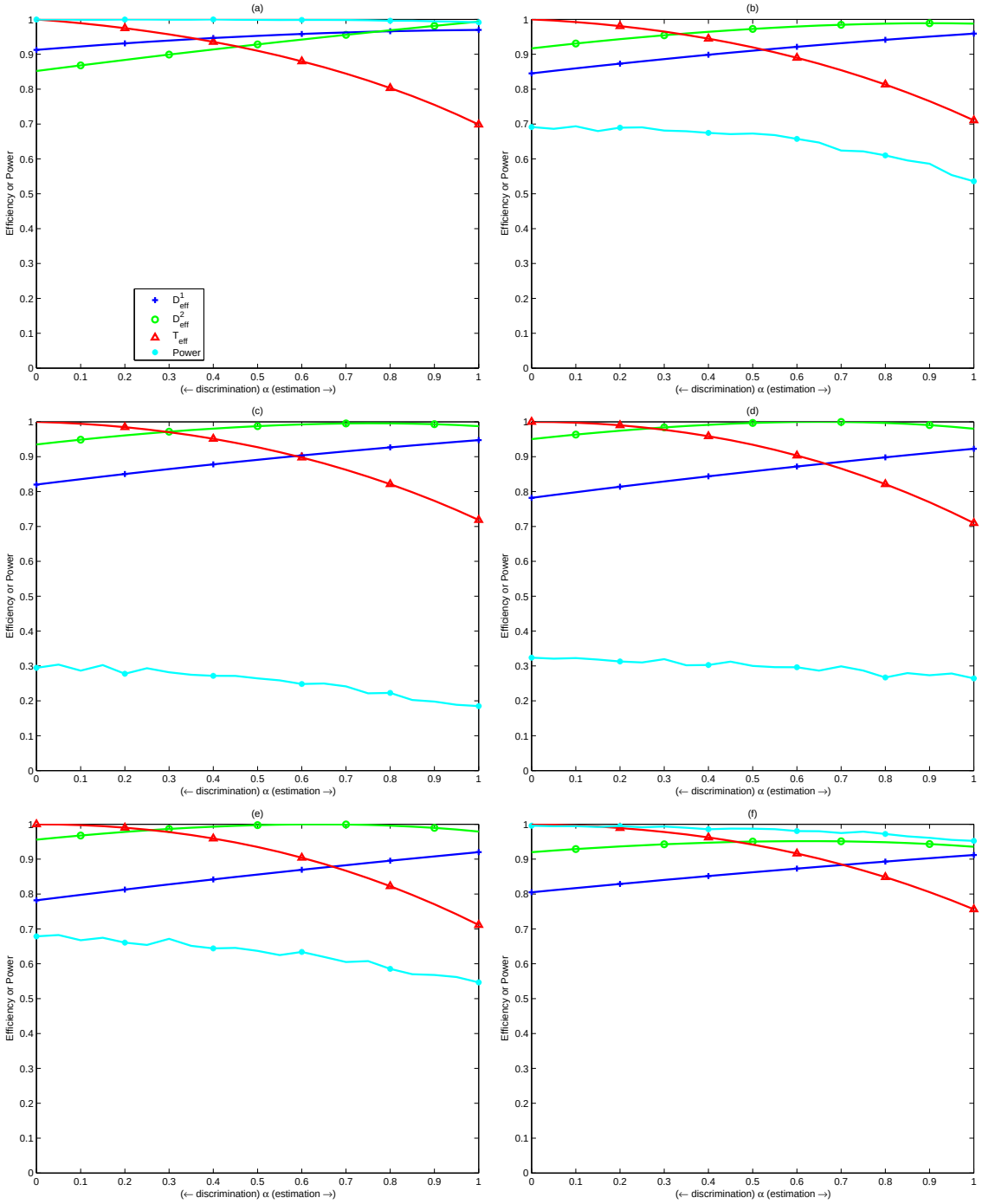

FIGURE 5.6: Efficiencies and power of hybrid designs for models M_1 and M_2, for (a) $\boldsymbol{\beta}_2 = (1, 1, 1, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 1, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 1, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 1, -1)'$. Corresponding values of $\hat{\boldsymbol{\beta}}_1$ for these values of $\boldsymbol{\beta}_2$ are given in Table 5.2.

Table 5.5: Product optimal designs for models M_1 and M_2 when $\beta_{22} = 1$.

$\hat{\boldsymbol{\beta}}_1$, $\boldsymbol{\beta}_2$		Product optimal design				
$\hat{\boldsymbol{\beta}}_1 = (0.7072, 0.7072, 0.7072)'$	x_1	−1.0000	−1.0000	0.1937	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 1, 1)'$	x_2	−1.0000	1.0000	1.0000	−1.0000	0.1938
	w	0.2757	0.2562	0.1060	0.2562	0.1060
$\hat{\boldsymbol{\beta}}_1 = (0.8679, 0.8679, 0.8679)'$	x_1	−1.0000	−1.0000	0.4676	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 1, 0.5)'$	x_2	−1.0000	1.0000	1.0000	−1.0000	0.4676
	w	0.2745	0.2655	0.0973	0.2655	0.0973
$\hat{\boldsymbol{\beta}}_1 = (0.9261, 0.9261, 0.9261)'$	x_1	−1.0000	−1.0000	0.8018	1.0000	
$\boldsymbol{\beta}_2 = (1, 1, 1, 0.3)'$	x_2	−1.0000	1.0000	0.8019	−1.0000	
	w	0.2746	0.2697	0.1861	0.2697	
$\hat{\boldsymbol{\beta}}_1 = (1.0523, 1.0523, 1.0523)'$	x_1	−1.0000	−1.0000	1.0000	1.0000	
$\boldsymbol{\beta}_2 = (1, 1, 1, -0.3)'$	x_2	−1.0000	1.0000	−1.0000	1.0000	
	w	0.2764	0.2764	0.2764	0.1707	
$\hat{\boldsymbol{\beta}}_1 = (1.0728, 1.0728, 1.0728)'$	x_1	−1.0000	−1.0000	1.0000	1.0000	
$\boldsymbol{\beta}_2 = (1, 1, 1, -0.5)'$	x_2	−1.0000	1.0000	−1.0000	1.0000	
	w	0.2771	0.2771	0.2771	0.1687	
$\hat{\boldsymbol{\beta}}_1 = (1.0707, 1.0706, 1.0706)'$	x_1	−1.0000	−1.0000	−0.9766	0.9777	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 1, -1)'$	x_2	−0.9745	0.9765	−1.0000	−1.0000	1.0000
	w	0.1169	0.2767	0.1592	0.2767	0.1705

5.3.5 Alternative parameter values

In the previous section, all designs were found assuming that the intercept and both main effect coefficients for the larger model were all 1 (i.e. that $\beta_{20} = \beta_{21} = \beta_{22} = 1$). In this section, all of the design optimisations are repeated with $\beta_{22} = 2$, that is it is supposed that the effect of the second factor x_2 is twice as large as the effect of the first factor x_1.

It can be seen from Figure 5.7 that again the region of interest for the interaction term is between −1 and 1. Hence, the design process is repeated for the same values of β_{23} as in the previous section. While the three corner points (−1,−1), (−1,1) and (1,−1) remain more or less fixed, the fourth support point is placed further from the corner as the value of the interaction coefficient increases.

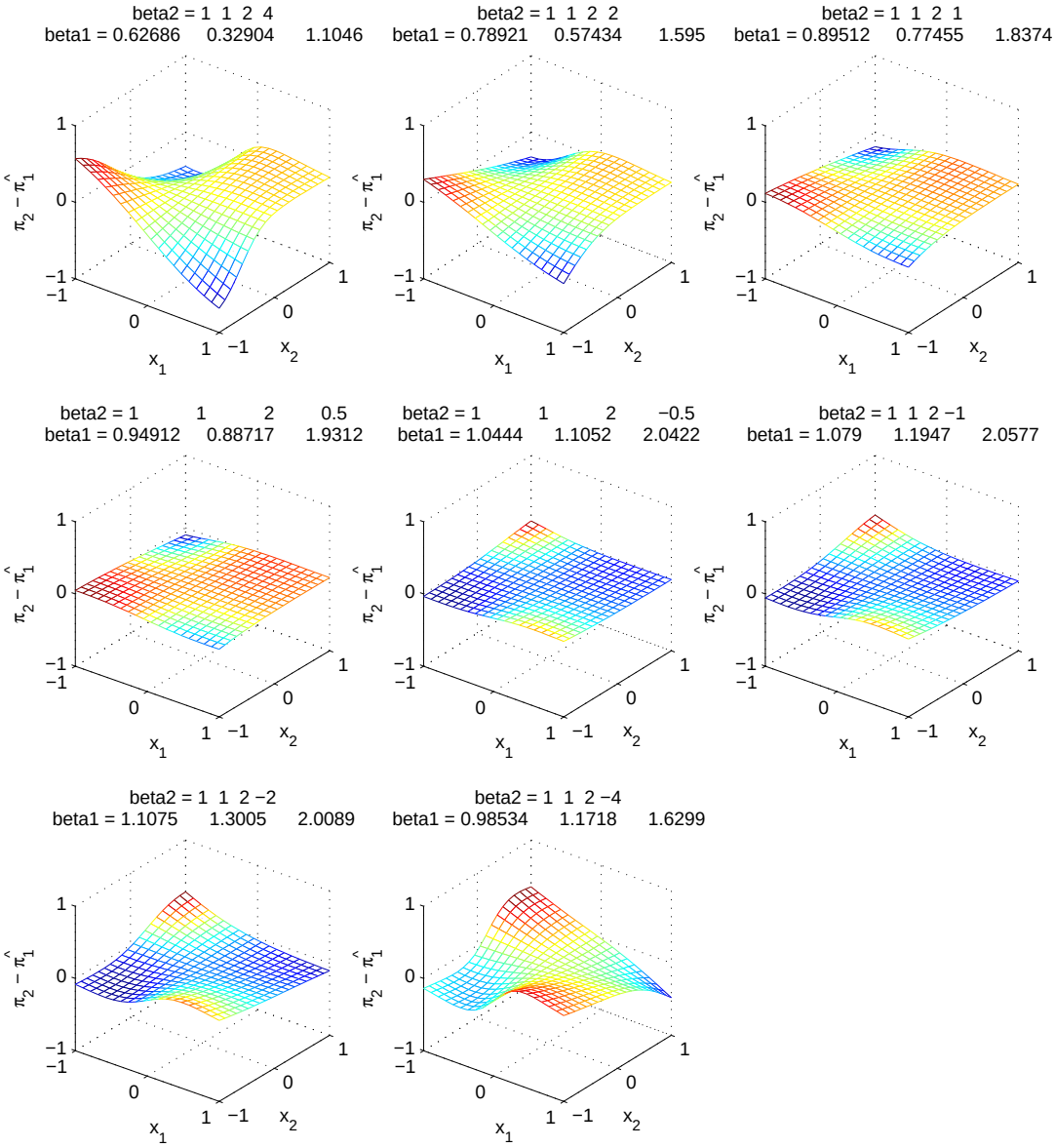

FIGURE 5.7: The difference between models M_1 and M_2 over various values of the interaction coefficient in M_2, when $\beta_{22} = 2$. For a given $\boldsymbol{\beta}_2$, $\hat{\boldsymbol{\beta}}_1$ is the best fit of M_1 to the expected responses under M_2. The colour of the surface is indicative of the difference between the expected response of the models: red signifies that π_2 is greater than $\hat{\pi}_1$, blue signifies that π_2 is less than $\hat{\pi}_1$.

Table 5.6: T-optimal designs for nested logistic models with two variables when $\beta_{22} = 2$.

$\boldsymbol{\beta}_2'$	T-optimal design				
$(1, 1, 2, 1)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	0.3947
	w	0.1944	0.1944	0.2285	0.3827
$(1, 1, 2, 0.5)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	0.4381
	w	0.1996	0.1996	0.2000	0.4007
$(1, 1, 2, 0.3)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	0.4708
	w	0.2016	0.2016	0.1885	0.4082
$(1, 1, 2, -0.3)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	0.6554
	w	0.2067	0.2067	0.1550	0.4316
$(1, 1, 2, -0.5)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-1.0000	1.0000	-1.0000	0.7620
	w	0.2082	0.2082	0.1441	0.4395
$(1, 1, 2, -1)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-0.9155	0.9155	-1.0000	1.0000
	w	0.2173	0.2173	0.1304	0.4350

The T-optimal designs for the various interaction parameter values are given in Table 5.6 and Figure 5.8. The fitted parameter values for model M_1 under the T-optimal designs are given in Table 5.7.

In contrast to the T-optimal designs for the original set of parameter values, the support points of these designs include some deviation from the four corners of the design region. Three corner points remain relatively fixed for all values of β_{23}, while the x_2 value of the 'top right' point (the point at $(1, 1)$ for $\beta_{23} = -1$) decreases as β_{23} increases.

D-optimal designs for models M_1 and M_2 are given in Tables 5.8 and 5.9, as well as in Figures 5.9 and 5.10. The product optimal designs are given in Table 5.10 and Figure 5.11.

The support points of the D-optimal designs for both models, as well as the product

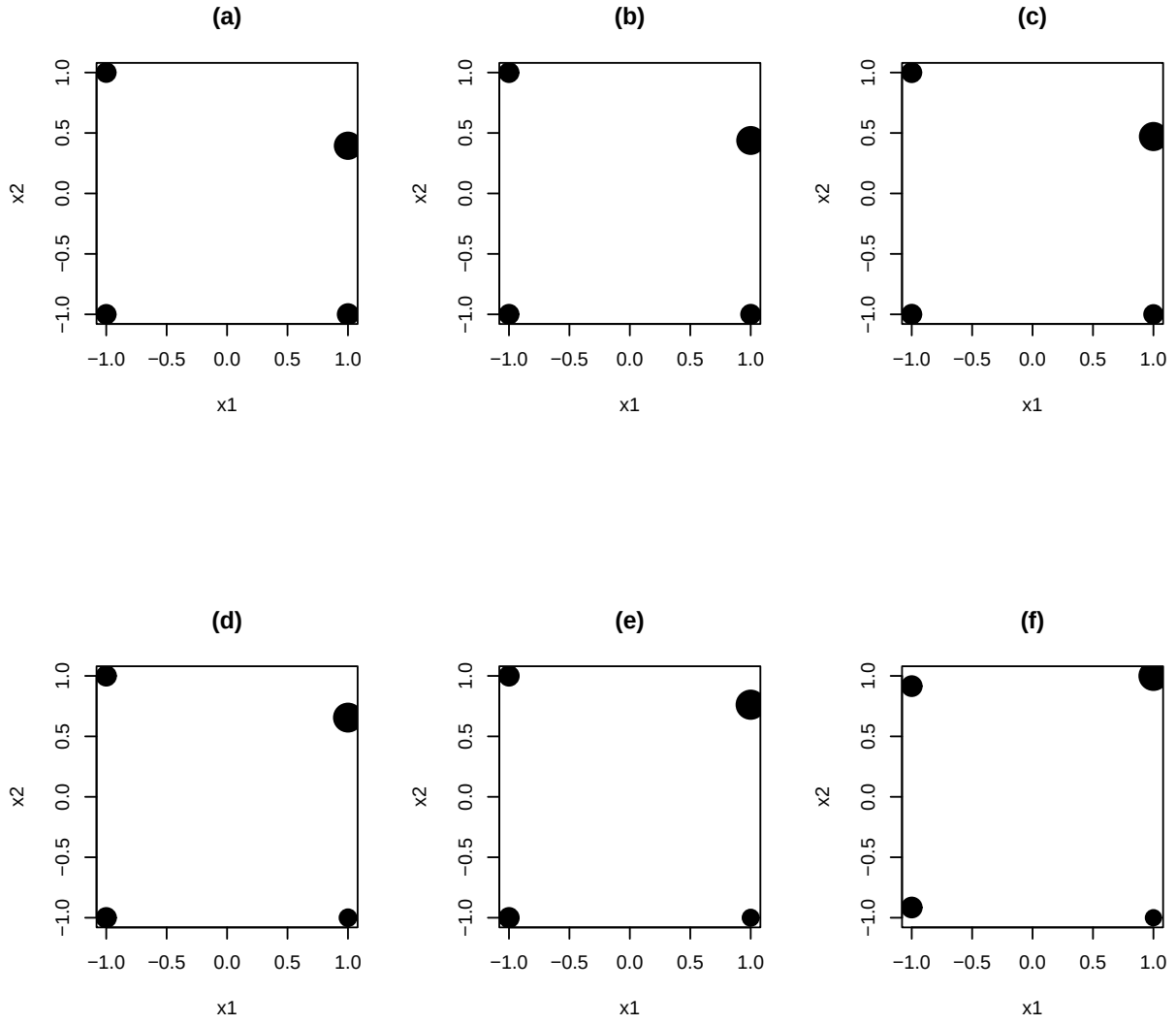

Figure 5.8: T-optimal designs for (a) $\boldsymbol{\beta}_2 = (1, 1, 2, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 2, -1)'$.

Table 5.7: Fitted values of the parameters of the smaller model, given by the T-optimal designs for the various values of $\boldsymbol{\beta}_2$, when $\beta_{22} = 2$.

$\boldsymbol{\beta}_2'$	$\hat{\boldsymbol{\beta}}_1'$
$(1, 1, 2, 1)$	$(0.6543, 0.6544, 1.6871)$
$(1, 1, 2, 0.5)$	$(0.8283, 0.8283, 1.8736)$
$(1, 1, 2, 0.3)$	$(0.8984, 0.8984, 1.9322)$
$(1, 1, 2, -0.3)$	$(1.0922, 1.0922, 2.0404)$
$(1, 1, 2, -0.5)$	$(1.1448, 1.1448, 2.0496)$
$(1, 1, 2, -1)$	$(1.2247, 1.2247, 1.9954)$

optimal designs, show a much greater deviation from the corners of the design region. This may be due to the increased nonlinearity of the response surface due to the increase of the coefficient of x_2. The difference between the D-optimal designs for the two models is less pronounced than for the original set of parameter values, although the designs for M_1 where $\beta_{23} < 0$ seem to require only three support points, whereas the corresponding designs for M_2 still include at least four points. These differences are reflected in the product optimal designs.

There are greater differences between the product optimal and T-optimal designs than was seen for the original set of parameter values. Given these differences, it may be reasonable to expect to see a greater difference in D- and T-efficiencies for the hybrid designs as α changes.

Hybrid designs are again calculated for a range of α, and their T-efficiencies, D-efficiencies and powers are calculated. The results for the various parameter values are shown in Figure 5.12.

It can again be seen that as alpha increases (reflecting an increasing importance placed on parameter estimation), the D-efficiencies increase, and the T-efficiency decreases.

While for the previous set of parameter values the cases $\beta_{23} = -1$ and $\beta_{23} = 1$ were not informative in terms of the power calculations due to the large difference between the models with and without the interaction term, the results for $\beta_{23} = -1$ are easier to interpret in this case. The powers of the designs calculated with $\beta_{23} = 1$ are still too high to be of any practical use in this investigation. A general (often very gentle) decrease in power can again

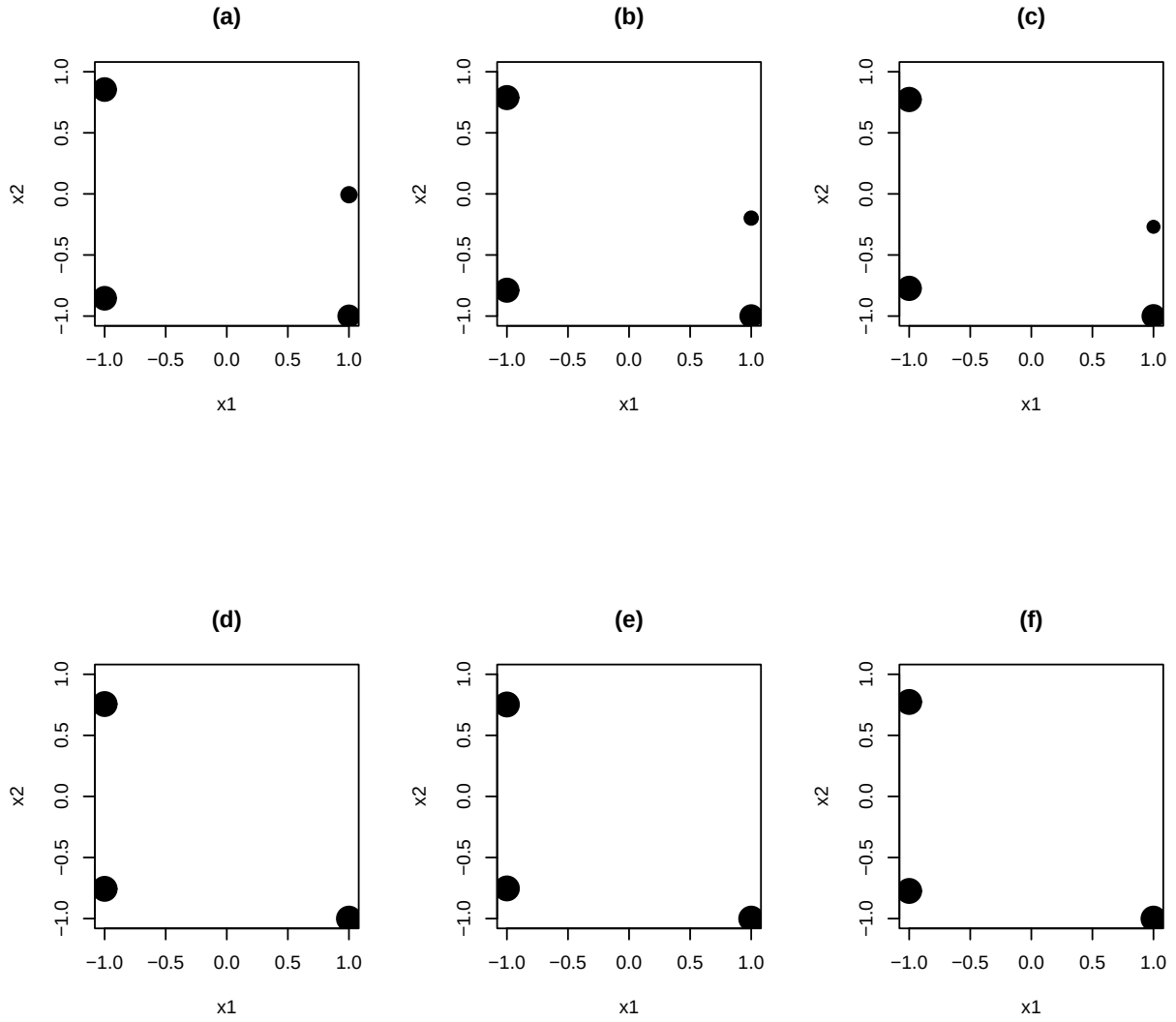

Figure 5.9: D-optimal designs for model M_1, for (a) $\boldsymbol{\beta}_2 = (0.6543, 0.6544, 1.6871)'$, (b) $\boldsymbol{\beta}_2 = (0.8283, 0.8283, 1.8736)'$, (c) $\boldsymbol{\beta}_2 = (0.8984, 0.8984, 1.9322)'$, (d) $\boldsymbol{\beta}_2 = (1.0922, 1.0922, 2.0404)'$, (e) $\boldsymbol{\beta}_2 = (1.1448, 1.1448, 2.0496)'$, and (f) $\boldsymbol{\beta}_2 = (1.2247, 1.2247, 1.9954)'$. The values of $\hat{\boldsymbol{\beta}}_1$ are taken from the T-optimal designs for the values of $\boldsymbol{\beta}_2$ given in Table 5.7.

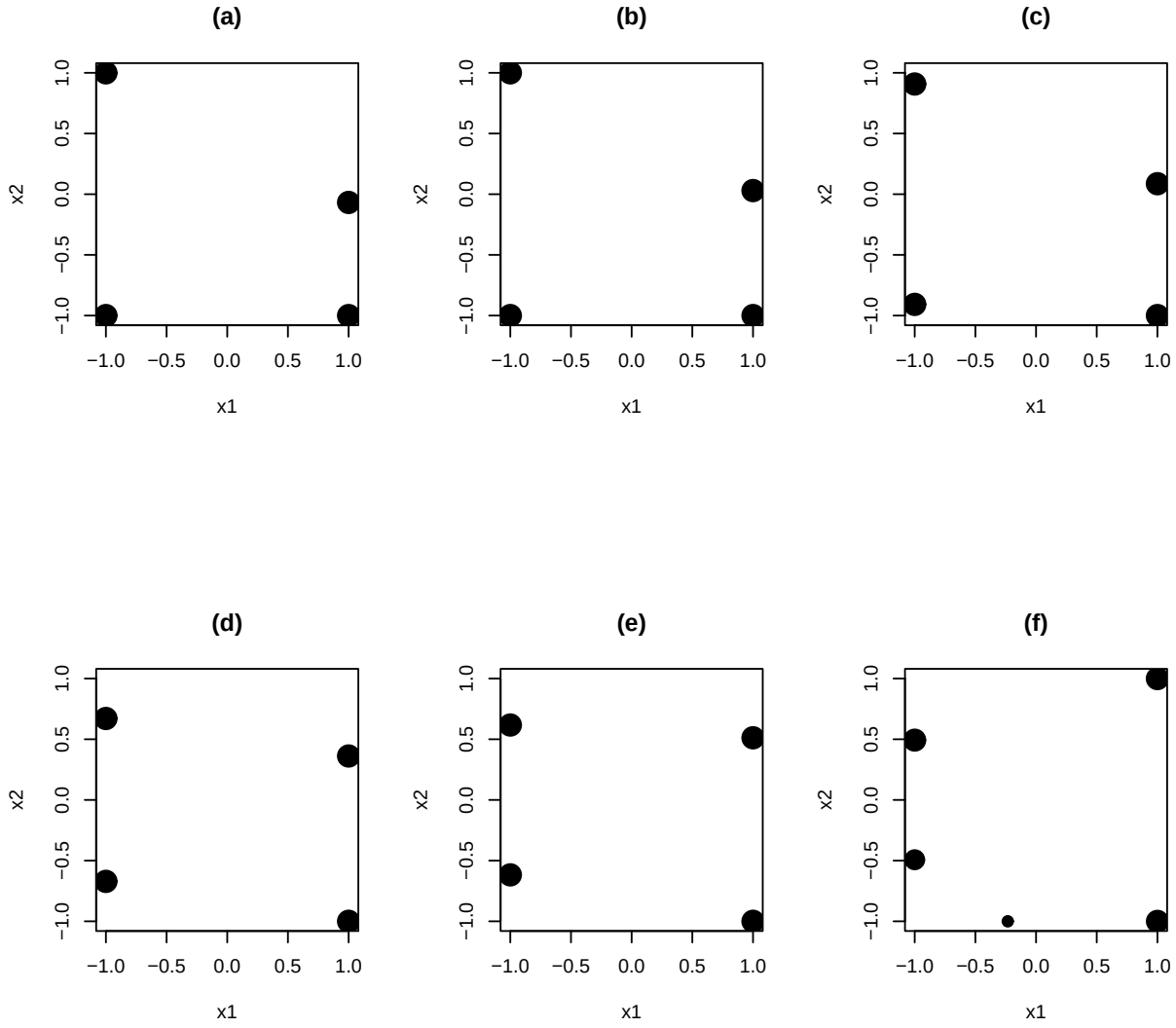

Figure 5.10: D-optimal designs for model M_2, for (a) $\boldsymbol{\beta}_2 = (1, 1, 2, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 2, -1)'$.

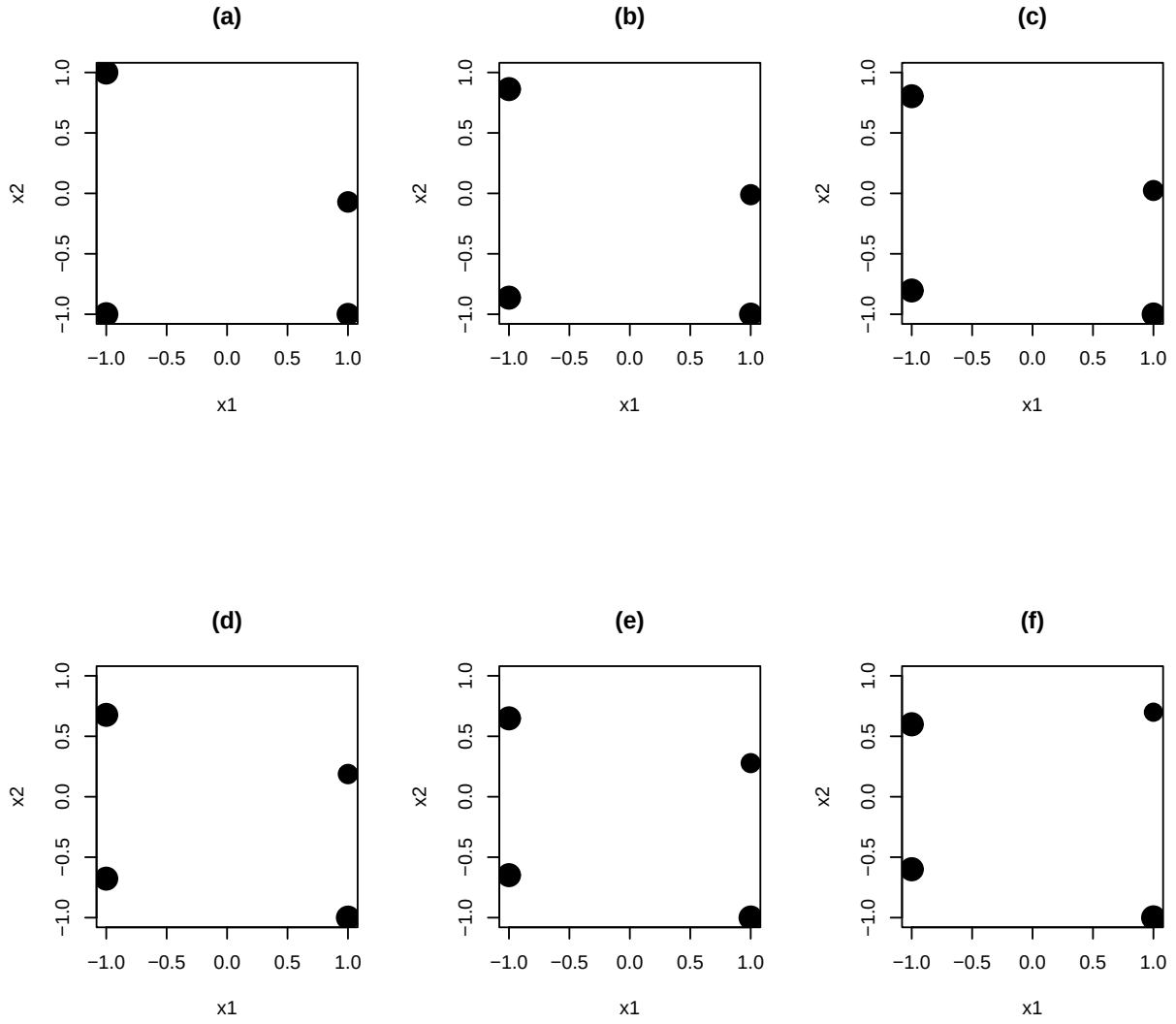

Figure 5.11: Product optimal designs for models M_1 and M_2, for (a) $\boldsymbol{\beta}_2 = (1, 1, 2, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 2, -1)'$. Corresponding values of $\hat{\boldsymbol{\beta}}_1$ for these values of $\boldsymbol{\beta}_2$ are given in Table 5.7.

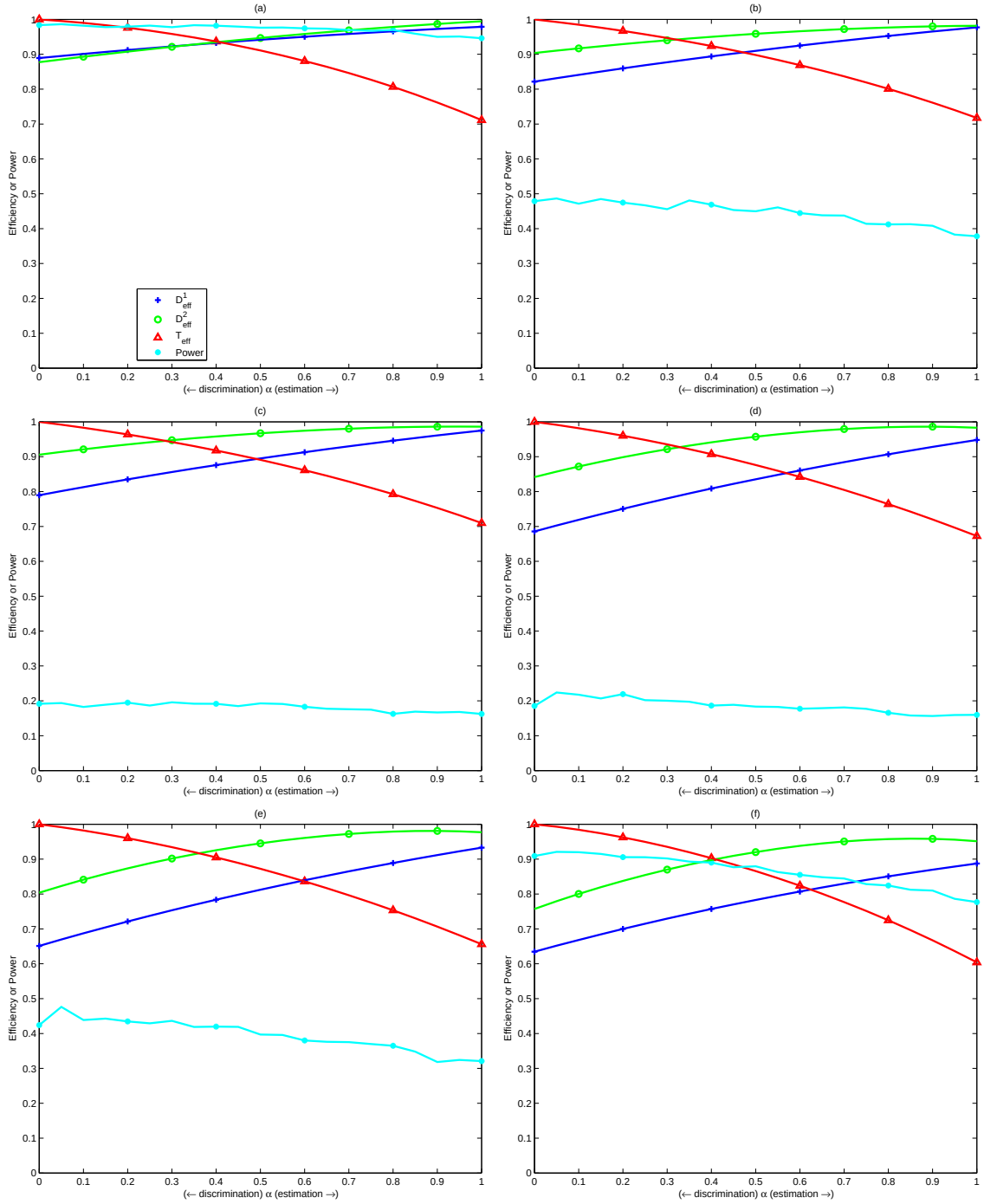

FIGURE 5.12: Efficiencies and power of hybrid designs for models M_1 and M_2, for (a) $\boldsymbol{\beta}_2 = (1, 1, 2, 1)'$, (b) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.5)'$, (c) $\boldsymbol{\beta}_2 = (1, 1, 2, 0.3)'$, (d) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.3)'$, (e) $\boldsymbol{\beta}_2 = (1, 1, 2, -0.5)'$, and (f) $\boldsymbol{\beta}_2 = (1, 1, 2, -1)'$. Corresponding values of $\hat{\boldsymbol{\beta}}_1$ for these values of $\boldsymbol{\beta}_2$ are given in Table 5.7.

Table 5.8: D-optimal designs for model M_1 when $\beta_{22} = 2$. The values of $\hat{\boldsymbol{\beta}}_1$ are taken from the T-optimal designs for the values of $\boldsymbol{\beta}_2$ given in Table 5.7.

$\hat{\boldsymbol{\beta}}_1'$		D-optimal design			
$(0.6543, 0.6544, 1.6871)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-0.8544	0.8544	-1.0000	-0.0067
	w	0.3003	0.3003	0.2612	0.1383
$(0.8283, 0.8283, 1.8736)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-0.7884	0.7884	-1.0000	-0.1980
	w	0.3110	0.3110	0.2730	0.1051
$(0.8984, 0.8984, 1.9322)$	x_1	-1.0000	-1.0000	1.0000	1.0000
	x_2	-0.7731	0.7730	-1.0000	-0.2691
	w	0.3162	0.3162	0.2816	0.0860
$(1.0922, 1.0922, 2.0404)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-0.7564	0.7564	-1.0000	
	w	0.3333	0.3333	0.3333	
$(1.1448, 1.1448, 2.0496)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-0.7530	0.7530	-1.0000	
	w	0.3333	0.3333	0.3333	
$(1.2247, 1.2247, 1.9954)$	x_1	-1.0000	-1.0000	1.0000	
	x_2	-0.7735	0.7735	-1.0000	
	w	0.3333	0.3333	0.3333	

be seen as α decreases, with the product design again showing only a marginally lower power than the T-optimal design in each case.

5.4 Discussion

The T-optimal and product optimal designs can easily (without further optimisation) be combined into a hybrid design which can address, to some extent, the separate objectives of model discrimination and parameter estimation, as seen in Chapter 3 for nonlinear models.

A hybrid design with more emphasis placed on parameter estimation (and hence more weight placed on the support points of the product optimal design) has greater D-efficiencies

Table 5.9: D-optimal designs for model M_2 when $\beta_{22} = 2$.

$\boldsymbol{\beta}_2'$		D-optimal design				
$(1, 1, 2, 1)$	x_1	-1.0000	-1.0000	1.0000	1.0000	
	x_2	1.0000	-1.0000	-0.0680	-1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 2, 0.5)$	x_1	-1.0000	-1.0000	1.0000	1.0000	
	x_2	1.0000	-1.0000	0.0298	-1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 2, 0.3)$	x_1	-1.0000	-1.0000	1.0000	1.0000	
	x_2	0.9079	-0.9079	0.0865	-1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 2, -0.3)$	x_1	-1.0000	-1.0000	1.0000	1.0000	
	x_2	0.6711	-0.6710	0.3621	-1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 2, -0.5)$	x_1	-1.0000	-1.0000	1.0000	1.0000	
	x_2	-0.6173	0.6174	0.5120	-1.0000	
	w	0.2500	0.2500	0.2500	0.2500	
$(1, 1, 2, -1)$	x_1	-1.0000	-1.0000	-0.2330	1.0000	1.0000
	x_2	-0.4928	0.4928	-1.0000	-1.0000	1.0000
	w	0.1990	0.2452	0.0611	0.2446	0.2500

but a lower T-efficiency (as defined here), and to some extent a lower power to discriminate between models. This difference in power, however, is generally minimal. The product optimal design may hence be an acceptable (and easy to calculate) design to address both discrimination and estimation objectives.

It appears that the T-efficiency of a design is a good indication as to whether the design will result in a high power to discriminate between the two models. The actual relationship between T-efficiency and power, however, is yet to be seen, and remains the topic of further research.

If the computational problems associated with the T_E-optimality criterion are overcome, it will be interesting to see whether designs constructed using this method have the same

Table 5.10: Product optimal designs for models M_1 and M_2 when $\beta_{22} = 2$.

$\hat{\boldsymbol{\beta}}_1$, $\boldsymbol{\beta}_2$		Product optimal design			
$\hat{\boldsymbol{\beta}}_1 = (0.6543, 0.6544, 1.6871)'$	x_1	-1.0000	-1.0000	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 2, 1)'$	x_2	-1.0000	1.0000	-1.0000	-0.0709
	w	0.2737	0.2737	0.2398	0.2128
$\hat{\boldsymbol{\beta}}_1 = (0.8283, 0.8283, 1.8736)'$	x_1	-1.0000	-1.0000	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 2, 0.5)'$	x_2	-0.8626	0.8626	-1.0000	-0.0108
	w	0.2705	0.2705	0.2531	0.2059
$\hat{\boldsymbol{\beta}}_1 = (0.8984, 0.8984, 1.9322)'$	x_1	-1.0000	-1.0000	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 2, 0.3)'$	x_2	-0.8035	0.8034	-1.0000	0.0238
	w	0.2694	0.2694	0.2586	0.2026
$\hat{\boldsymbol{\beta}}_1 = (1.0922, 1.0922, 2.0404)'$	x_1	-1.0000	-1.0000	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 2, -0.3)'$	x_2	-0.6773	0.6773	-1.0000	0.1878
	w	0.2688	0.2688	0.2728	0.1896
$\hat{\boldsymbol{\beta}}_1 = (1.1448, 1.1448, 2.0496)'$	x_1	-1.0000	-1.0000	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 2, -0.5)'$	x_2	-0.6485	0.6485	-1.0000	0.2782
	w	0.2696	0.2696	0.2768	0.1841
$\hat{\boldsymbol{\beta}}_1 = (1.2247, 1.2247, 1.9954)'$	x_1	-1.0000	-1.0000	1.0000	1.0000
$\boldsymbol{\beta}_2 = (1, 1, 2, -1)'$	x_2	-0.5996	0.5995	-1.0000	0.6993
	w	0.2741	0.2741	0.2854	0.1664

issues as those described in this chapter for T-optimal designs.

Chapter 6

Optimal crossover designs for generalised linear models

To complement the pharmacological application of optimal design for nonlinear models in Chapter 4, this chapter is motivated by an application of optimal design for generalised linear models to a hypothetical pharmacodynamic experiment, where crossover designs are frequently used.

Pharmacodynamics (PD) is the study of the biochemical and physiological effects of drugs. The construction of optimal designs for dose-ranging trials with multiple periods is considered in this chapter, where the outcome of the trial (the effect of the drug) is considered to be a binary response: the success or failure of a drug to bring about a particular change in the patient after a given amount of time. The carryover effect of each dose from one period to the next is assumed to be proportional to the direct effect. It is shown for a logistic regression model that the efficiency of optimal parallel (single-period) or crossover (two-period) design is substantially greater than a balanced design. The optimal designs are also shown to be robust to misspecification of the value of the parameters. Finally, the parallel and crossover designs are combined to provide the experimenter with greater flexibility.

6.1 Introduction

The drug development process has seen a number of changes over the past 30 years. Many of these changes that are pertinent to the design of studies have been oriented to the development of more efficient design processes. From a pharmacological perspective, a major improvement in efficiency has been afforded by the integration of advanced statistical methodology into pharmacokinetic studies (Aarons, 1999). Traditionally population pharmacokinetic studies were based on intensive sampling strategies, where the same number of blood samples (often greater than 12) were taken from each patient. Since the introduction of non-linear mixed effects modelling in the early 1980s (see for example Sheiner and Beal (1980)), unbalanced, 'sparse' designs have become common practice in population pharmacokinetic studies. For example, the number and timing of blood samples can vary from subject to subject, with some subjects providing fewer samples than parameters to estimate. In contrast to the flexibility now afforded to population pharmacokinetic (PK) studies, the design of crossover trials with multiple treatment periods and treatment sequences are typically limited to balanced designs, where all individuals receive all treatments and block sequences are designed so that all treatments follow all other treatments at some stage. This produces a design which is balanced with respect to carryover effects. Recent research into the design of crossover trials for PK studies can be found in Jones and Wang (1998), Jones *et al.* (1999) and Jones and Wang (1999). They show, for two treatments only, that balanced designs are optimal for trials that have no carryover effect. This work is theoretical, with no presentation of practical application. It is likely that sparse period and unbalanced designs will provide an efficient means in which to test the efficacy of a drug without the need for time consuming and expensive balanced designs. The aim of this paper is to assess different approaches to optimising treatment allocation in parallel and crossover designed studies.

This chapter first introduces the model for a hypothetical drug with a binary response, and gives the information matrix for a crossover design with two periods (which can be simplified for the single period case). Optimal and balanced parallel designs (where each subject receives only one dose) are then compared in terms of their efficiency. Another comparison of optimal and balanced allocations is given, this time for crossover designs,

where two doses are given to each subject. This is considered for two cases: where the amount of carryover effect from one period to the next is known, and again when it is unknown. The sensitivity of the efficiency of these designs to parameter misspecification is then assessed. Finally, composites of both optimal parallel and crossover designs are given, and their efficiencies are evaluated.

6.2 Model

Consider a fictive drug, which elicits an all or nothing response: success or failure. For the purposes of this study the fictive drug is based on the triptan like 5-HT_{1D} agonists for the treatment of migraine (see for example Nestorov *et al.* (2001)). In these scenarios it is assumed that the patients' response can be described by a binary outcome defined by treatment success (1) or failure (0) corresponding to alleviation or persistence of the migraine headache at some predefined time after the dose, respectively. We have assumed that the dose levels may be allowed to vary between 0 and 20 units, each with a fixed probability of producing a success, dependent on the size of the dose(s) given. We aim to find an optimum allocation of dosage sequences (with a maximum of 2 doses per subject), and compare it to a balanced allocation. The variation in the data arising from a crossover trial is usually explained by an analysis-of-variance model such as

$$Y_{kl} = \beta + \theta_{d[k,l]} + \lambda_{d[k,l-1]} + \rho_l + s_k + \epsilon_{kl},$$

where Y_{kl} is the response of subject k in period l to dose (treatment) $d[k,l]$. $\theta_{d[k,l]}$ is the effect of dose $d[k,l]$, $\lambda_{d[k,l-1]}$ is the carryover effect of the dose given in period $l-1$, ρ_l is the effect of period l, s_k is the effect of the kth subject, β is the overall mean and ϵ_{kl} is the error term. It can be convenient to consider the carryover effect of a dose as directly proportional to its direct effect (Kempton *et al.*, 2001), in which case $\lambda_{d[k,l-1]}$ would be written as $\alpha\theta_{d[k,l-1]}$. Another term that may be included in such a model is the interaction between subject and period effects.

In the example used in this chapter, the subject and period effects are ignored. A number of references (including Robinson and Jewell (1991) and Yano *et al.* (2001)) support

this assumption by showing that the inclusion of subject effects do not necessarily yield more precise estimates of the parameters of primary interest in logistic regression. However, the inclusion of random subject effects in this type of model is the focus of the following chapter, where mixed effects models are considered.

The response y_{ij} of the jth patient to the ith dose is modelled as a logistic regression model with $p = 3$ parameters,

$$\begin{aligned} \text{logit}\{\pi_{ij}\} &= \text{logit}\{P(y_{ij} = 1)\} = \theta_0 + \theta_1 d_{ij} + \alpha\theta_1 d_{(i-1)j} \\ & \qquad i = 1, 2, \quad j = 1, \ldots, N, \end{aligned} \tag{6.1}$$

where for $i \geq 1$, d_{ij} is the size of the ith dose given to the jth patient, and $d_{0j} = 0$. In this model, the residual effect of a particular dose is considered to be directly proportional to its direct effect (see Kempton *et al.* (2001)). If α, the level of proportional residual effect, is assumed to be known, the vector of model parameters is $\boldsymbol{\theta} = (\theta_0, \theta_1)'$. In this case, the model is a generalised linear model (GLM), where the predictor (the right hand side of Equation (6.1)) is linear in the model parameters. This is assumed in Section 6.4.1, whereas in Section 6.4.2 α is considered to be unknown, and so $\boldsymbol{\theta} = (\theta_0, \theta_1, \alpha)'$, in which case the right hand side of Equation (6.1) becomes nonlinear in the parameters, and is no longer strictly a GLM. In either case, the probabilities are given by

$$\begin{aligned} \pi_{ij} &= \text{logit}^{-1}\{\theta_0 + \theta_1 d_{ij} + \alpha\theta_1 d_{(i-1)j}\} \\ &= \frac{\exp\{\theta_0 + \theta_1 d_{ij} + \alpha\theta_1 d_{(i-1)j}\}}{1 + \exp\{\theta_0 + \theta_1 d_{ij} + \alpha\theta_1 d_{(i-1)j}\}}. \end{aligned}$$

Although the carry over effect can be assumed to be negligible for pharmacokinetic studies if a sufficient washout period is included (as per Senn and Ezzet (1999)), the same assumption is not necessarily valid for pharmacodynamic studies.

The response of the jth subject after a single dose (with no previous doses) may be simplified to

$$\text{logit}\{\pi_{1j}\} = \log\left(\frac{\pi_{1j}}{1 - \pi_{1j}}\right) = \theta_0 + \theta_1 d_{1j}, \tag{6.2}$$

where similarly π_{1j} is the probability of success of dose d_{1j} alone, given by

$$\begin{aligned} \pi_{1j} &= \text{logit}^{-1}\{\theta_0 + \theta_1 d_{1j}\} \\ &= \frac{\exp\{\theta_0 + \theta_1 d_{1j}\}}{1 + \exp\{\theta_0 + \theta_1 d_{1j}\}}. \end{aligned}$$

For an n-point continuous design (as described in Section 2.1), the log-likelihood function for the responses in the ith period ($i = 1, 2$) is given by

$$\ell(\boldsymbol{\pi}_i, \boldsymbol{y}_i) = \sum_{j=1}^{n} w_j \log \left\{ \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \right\},$$

where w_j are the design weights, $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{in})'$ and $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in})'$.

6.3 Parallel designs

Optimal and balanced parallel designs

Parallel designs are those with only one dose per subject and so the model is simply the logistic model given in Equation (6.2), as there are only single doses. The support point of the jth elementary design is hence $\boldsymbol{\xi}_j = d_{1j}$. We have a standard logistic regression model with $\boldsymbol{\theta} = (\theta_0, \theta_1)'$, and so the information matrix is as given in Equation (2.7):

$$\boldsymbol{M}_1(\boldsymbol{\theta}, \xi) = \boldsymbol{X}_1' \boldsymbol{W}_1 \boldsymbol{X}_1 \tag{6.3}$$

where

$$\boldsymbol{X}_1 = \begin{bmatrix} 1 & d_{11} \\ \vdots & \vdots \\ 1 & d_{1n} \end{bmatrix}$$

and $\boldsymbol{W}_1 = \text{diag}\{w_j \pi_{1j}(1 - \pi_{1j})\}$.

The true values of θ_0 and θ_1 are unknown, but are necessary for the calculation of $\boldsymbol{M}_1$. The values of these parameters may be chosen (estimated) from past experience or based on expert opinion. For the purposes of this investigation select $\theta_0 = -1$ and $\theta_1 = 0.3$, i.e. use the parameter vector $\boldsymbol{\theta}^0 = (-1, 0.3)'$. These parameter values yield responses similar to other triptans where a positive response is seen in approximately 30% of patients for a zero dose, and approximately 70% of patients respond positively to the maximum dose (20 units in this case) (Nestorov *et al.*, 2001). Consideration is later given to the sensitivity of the optimal design to this choice of parameter values.

The optimal parallel design ξ^{*p} ('p' to denote 'parallel') was found using the simulated annealing algorithm described in Section 2.4.1:

$$\xi^{*p} = \begin{Bmatrix} 0.00 & 9.32 \\ 0.5 & 0.5 \end{Bmatrix}.$$

The most likely practical application of this optimal design is to assign half of the patients to a group receiving the zero dose, and the remaining patients to a group receiving 10 units of the dose. This deviation from the optimal design is still 99.5% as efficient as the original design. This optimal design may be compared to a balanced parallel design:

$$\xi^{bp1} = \begin{Bmatrix} 0 & 5 & 10 & 20 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{Bmatrix}.$$

This balanced design has an efficiency of 0.7571 compared to the optimal design. This suggests that about 32% more subjects would be required for the balanced design to provide the same information as the optimal design. An alternative balanced design with four dose levels may be defined with a linear dose range:

$$\xi^{bp2} = \begin{Bmatrix} 0 & 5 & 10 & 15 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{Bmatrix}.$$

This design has an efficiency of 0.8367, meaning that approximately 20% more subjects would be required to match the information provided by the optimal design. A further alternative still could be the balanced parallel design with five dose levels:

$$\xi^{bp3} = \begin{Bmatrix} 0 & 5 & 10 & 15 & 20 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{Bmatrix}.$$

This design with more support points actually has a lower efficiency (0.7217) than the 4-point alternatives.

6.4 Cross-over designs

6.4.1 Optimal and balanced crossover designs, α known

Now consider crossover designs as an alternative to parallel designs, where each subject receives two doses, and the carryover effect α is assumed to be known. This case is quite

artificial and is included as a comparator for the parallel design, where the carryover effect is of no consequence. The elementary design for the j support point is now $\boldsymbol{\xi}_j = (d_{1j}, d_{2j})'$. As discussed in Section 6.2, when α is treated as a known fixed constant, the model is the standard GLM with the logit link function. The information matrix is therefore given by

$$\boldsymbol{M}(\boldsymbol{\theta}, \xi) = \boldsymbol{M}_1(\boldsymbol{\theta}, \xi) + \boldsymbol{M}_2(\boldsymbol{\theta}, \xi),$$

where $\boldsymbol{M}_1$ is the information matrix for the first period responses, given in Equation (6.3). $\boldsymbol{M}_2$ is the information matrix for the second period, given by

$$\boldsymbol{M}_2(\boldsymbol{\theta}, \xi) = \boldsymbol{X}_2' \boldsymbol{W}_2 \boldsymbol{X}_2 \tag{6.4}$$

where

$$\boldsymbol{X}_2 = \begin{bmatrix} 1 & d_{21} + \alpha d_{11} \\ \vdots & \vdots \\ 1 & d_{2n} + \alpha d_{1n} \end{bmatrix}$$

and $\boldsymbol{W}_2 = \text{diag}\{w_i \pi_{2i}(1 - \pi_{2i})\}$.

Since $\boldsymbol{M}$ is again dependent on $\theta = (\theta_0, \theta_1)'$, as well as α, it is necessary to specify or know their values in order to evaluate the optimal design. The values of θ_0 and θ_1 are again set at -1 and 0.3, respectively, to give a parameter vector of $\boldsymbol{\theta}^0 = (-1, 0.3)'$, and we choose a proportional carryover factor of $\alpha = 0.25$. The degree of sensitivity of the optimal design to these parameter values is investigated later.

The optimal design in this case, ξ^{*ck}, where 'ck' denotes '<u>c</u>ross-over with α <u>k</u>nown', is

$$\xi^{*ck} = \left\{ \begin{matrix} (0.00, 9.32) \\ 1 \end{matrix} \right\},$$

where the ordered pair in the first row, as for all crossover designs presented from here on, represents the pair of dose levels (d_{1j}, d_{2j}), and the second row contains the design weights. That is, in this case the design requires that all subjects receive a 0 unit dose (placebo) in the first period, followed by a 9.32 unit dose in the second period. A likely practical application of this design would require that all patients receive 0 units followed by 10 units of the drug. This design is 99.5% as efficient as the optimal design.

Table 6.1: First balanced crossover design with four dose levels, ξ^{bc1}. The support points of the design are given by the combinations of the levels of the first and second doses, the design weights are given by the corresponding table entries.

		Second dose			
		0	5	10	20
First dose	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	5	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	10	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	20	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Table 6.2: Second balanced crossover design with four dose levels, ξ^{bc2}. The support points of the design are given by the combinations of the levels of the first and second doses, the design weights are given by the corresponding table entries.

		Second dose			
		0	5	10	15
First dose	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	5	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	10	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
	15	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

This design is compared to two potential balanced crossover designs with four dose levels, as shown in Tables 6.1 and 6.2. The first balanced crossover design, ξ^{bc1}, with four potential dose levels 0, 5, 10 and 20 units, has an efficiency of 0.7087 compared to the optimal design ξ^{*ck}. This balanced design would require approximately 41% more subjects to gain the same information as the optimal design. The other balanced crossover design, ξ^{bc2}, consists of dose levels 0, 5, 10 and 15 units, and has an efficiency of 0.7819. This means that 28% more subjects would be required to match the optimal design in terms of information gained from the study.

6.4.2 Optimal and balanced crossover designs, α unknown

Here it is assumed that α is now an unknown parameter (a more likely situation), that is the carryover effect, usually a nuisance parameter, is now included in the design process. The

information matrix again consists of the sum of the information from both periods:

$$\boldsymbol{M}(\boldsymbol{\theta}, \xi) = \boldsymbol{M}_1(\boldsymbol{\theta}, \xi) + \boldsymbol{M}_2(\boldsymbol{\theta}, \xi),$$

but the matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ now each incorporate a row and column corresponding to the new parameter α. The situation is now complicated slightly, as the model for the second period response is no longer strictly a GLM since the predictor

$$\eta_{2j} = \theta_0 + \theta_1 d_{2j} + \alpha\theta_1 d_{1j}$$

is not linear in the parameters.

Consider the log-likelihood function for the responses from period i:

$$\ell(\boldsymbol{\pi}_i, \boldsymbol{y}_i) = \sum_{j=1}^{n} w_j \log\left\{\pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}}\right\},$$

where w_j are the design weights. Taking the second derivative of this function with respect to any two parameters θ_u and θ_v yields

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\pi}_i, \boldsymbol{y}_i)}{\partial\theta_u \partial\theta_v} &= \sum_{j=1}^{n}\left[\frac{y_{ij}}{\pi_{ij}}\frac{\partial^2\pi_{ij}}{\partial\theta_u\partial\theta_v} - \frac{y_{ij}}{\pi_{ij}^2}\frac{\partial\pi_{ij}}{\partial\theta_u}\frac{\partial\pi_{ij}}{\partial\theta_v}\right] - \\ &\quad - \sum_{j=1}^{n}\left[\frac{1-y_{ij}}{1-\pi_{ij}}\frac{\partial^2\pi_{ij}}{\partial\theta_u\partial\theta_v} + \frac{1-y_{ij}}{(1-\pi_{ij})^2}\frac{\partial\pi_{ij}}{\partial\theta_u}\frac{\partial\pi_{ij}}{\partial\theta_v}\right].\end{aligned}$$

Now, as $\mathrm{E}(y_{ij}) = \pi_{ij}$ and $\frac{\partial\pi_{ij}}{\partial\theta_u} = \pi_{ij}(1-\pi_{ij})\frac{\partial\eta_{ij}}{\partial\theta_u}$, we have

$$\boldsymbol{M}_i(\boldsymbol{\theta}, \xi) = \mathrm{E}\left[-\frac{\partial^2\ell(\boldsymbol{\pi}_i, \boldsymbol{y}_i)}{\partial\theta_u\partial\theta_v}\right] = \sum_{j=1}^{n}\frac{\partial\eta_{ij}}{\partial\theta_u}\frac{\partial\eta_{ij}}{\partial\theta_v}\pi_{ij}(1-\pi_{ij}) = \boldsymbol{F}_i'\boldsymbol{W}_i\boldsymbol{F}_i,$$

where

$$\boldsymbol{F}_i = \begin{bmatrix} \frac{\partial\eta_{i1}}{\partial\theta_1} & \cdots & \frac{\partial\eta_{i1}}{\partial\theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial\eta_{in}}{\partial\theta_1} & \cdots & \frac{\partial\eta_{in}}{\partial\theta_p} \end{bmatrix}$$

and $\boldsymbol{W}_i = \mathrm{diag}\{w_j\pi_{ij}(1-\pi_{ij})\}$, as defined previously. Specifically, we have

$$\boldsymbol{F}_1 = \begin{bmatrix} 1 & d_{11} & 0 \\ \vdots & \vdots & \vdots \\ 1 & d_{1n} & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{F}_2 = \begin{bmatrix} 1 & d_{21}+\alpha d_{11} & \theta_1 d_{11} \\ \vdots & \vdots & \vdots \\ 1 & d_{2n}+\alpha d_{1n} & \theta_1 d_{1n} \end{bmatrix}.$$

The D-optimal design ξ^{*cu} below, where 'cu' denotes '<u>c</u>ross-over with α <u>u</u>nknown', was found by using this information matrix and with parameter vector $\boldsymbol{\theta}^0 = (-1, 0.3, 0.25)'$:

$$\xi^{*cu} = \left\{ \begin{matrix} (0.00, 0.00) & (0.00, 9.07) & (20.00, 0.00) \\ 0.039 & 0.572 & 0.389 \end{matrix} \right\}.$$

A more practical version of this design might be the following:

$$\xi^{*cu} = \left\{ \begin{matrix} (0, 0) & (0, 10) & (20, 0) \\ 0.05 & 0.55 & 0.4 \end{matrix} \right\}.$$

This design is 99.5% as efficient as ξ^{*cu}. The optimal design was again compared to the two potential balanced crossover designs given in Tables 6.1 and 6.2, as used for the previous case of α known. The first balanced crossover design, ξ^{bc1} has an efficiency of 0.6581 compared to this optimal design ξ^{*cu}. This balanced design would require approximately 52% more subjects to gain the same information as the optimal design. The other balanced crossover design, ξ^{bc2} has a slightly lower efficiency of 0.6394, meaning that 56% more subjects would be required to match the optimal design in terms of the information gained from the study. The higher efficiency and practicality of the optimal design, with only three combinations of dose levels, makes it a much more attractive alternative to the experimenter.

6.5 Sensitivity analysis

All D-optimal designs described in this chapter are locally optimal, that is they depend on a particular set of parameter values. This section investigates the impact that this has if the true parameter values vary from the specified values used to generate the designs.

A 'joint' sensitivity analysis is considered here, where the impact of varying all parameters together is considered, rather than a 'marginal' sensitivity analysis which would involve investigating the effects of varying only one parameter at a time.

For the purpose of this analysis, the dependence of the optimal designs on the parameter values is emphasised by writing the design as a function of $\boldsymbol{\theta}$: $\xi^*(\boldsymbol{\theta})$. The original optimal designs were calculated using $\boldsymbol{\theta}^0 = (-1, 0.3, 0.25)'$ (the last element of $\boldsymbol{\theta}^0$ was of course only included for the final optimal design), so we call these designs $\xi^{*p}(\boldsymbol{\theta}^0)$, $\xi^{*ck}(\boldsymbol{\theta}^0)$ and $\xi^{*cu}(\boldsymbol{\theta}^0)$.

The sensitivity analysis involves generating a number (500 in this case) of new parameter vectors by sampling uniformly between $(-1.5, 0.1, 0.1)'$ and $(-0.5, 0.5, 0.4)'$. This ensures a reasonable spread of parameter values, centred on $\boldsymbol{\theta}^0$. For each new parameter vector $\boldsymbol{\theta}_i$ $(i = 1, \ldots, 500)$, the three optimal designs $\xi^{*p}(\boldsymbol{\theta}_i)$, $\xi^{*ck}(\boldsymbol{\theta}_i)$ and $\xi^{*cu}(\boldsymbol{\theta}_i)$ are generated and compared to the original optimal designs by calculating the following efficiencies:

$$\begin{aligned}
\text{Eff}_i^p &= \frac{|\boldsymbol{M}(\boldsymbol{\theta}_i, \xi^{*p}(\boldsymbol{\theta}^0))|^{1/2}}{|\boldsymbol{M}(\boldsymbol{\theta}_i, \xi^{*p}(\boldsymbol{\theta}_i))|^{1/2}} \\
\text{Eff}_i^{ck} &= \frac{|\boldsymbol{M}(\boldsymbol{\theta}_i, \xi^{*ck}(\boldsymbol{\theta}^0))|^{1/2}}{|\boldsymbol{M}(\boldsymbol{\theta}_i, \xi^{*ck}(\boldsymbol{\theta}_i))|^{1/2}} \\
\text{Eff}_i^{cu} &= \frac{|\boldsymbol{M}(\boldsymbol{\theta}_i, \xi^{*cu}(\boldsymbol{\theta}^0))|^{1/3}}{|\boldsymbol{M}(\boldsymbol{\theta}_i, \xi^{*cu}(\boldsymbol{\theta}_i))|^{1/3}}, \quad i = 1, \ldots, 500
\end{aligned}$$

These efficiencies are represented by the histograms in Figure 6.1. It can be seen from the histograms that the efficiencies of the original optimal designs rarely fall below 0.7 in each case, and are above 0.85 for at least half of the simulated parameter values in each case. The optimal design for the case where α is unknown is particularly robust, with an efficiency of at 0.85 or greater for a very large number of new parameter values.

For comparison, the histograms of efficiencies of the corresponding balanced designs are also shown. The balanced designs with dose levels 0, 5, 10 and 15 units (ξ^{bp2} and ξ^{bc2}) were used for a fair comparison, as they were usually the most efficient type of balanced design (being only slightly less efficient for the case where α is unknown). The efficiency corresponding to the original parameter values (the mid-point of the ranges used in these simulations) are also shown by the dotted lines. For each of the three types of designs, the balanced design is more likely than not to be less efficient for different parameter values than it is for the originally specified parameter values. In contrast to the optimal designs, the efficiencies of the balanced designs rarely rise above 0.85, and often fall below 0.7. The efficiency of the balanced design when α is unknown is particularly poor.

6.6 Union designs

Finally, consider a union of parallel and crossover designs, in which some subjects receive one dose and some receive two. Greater information will obviously be gained from the patients

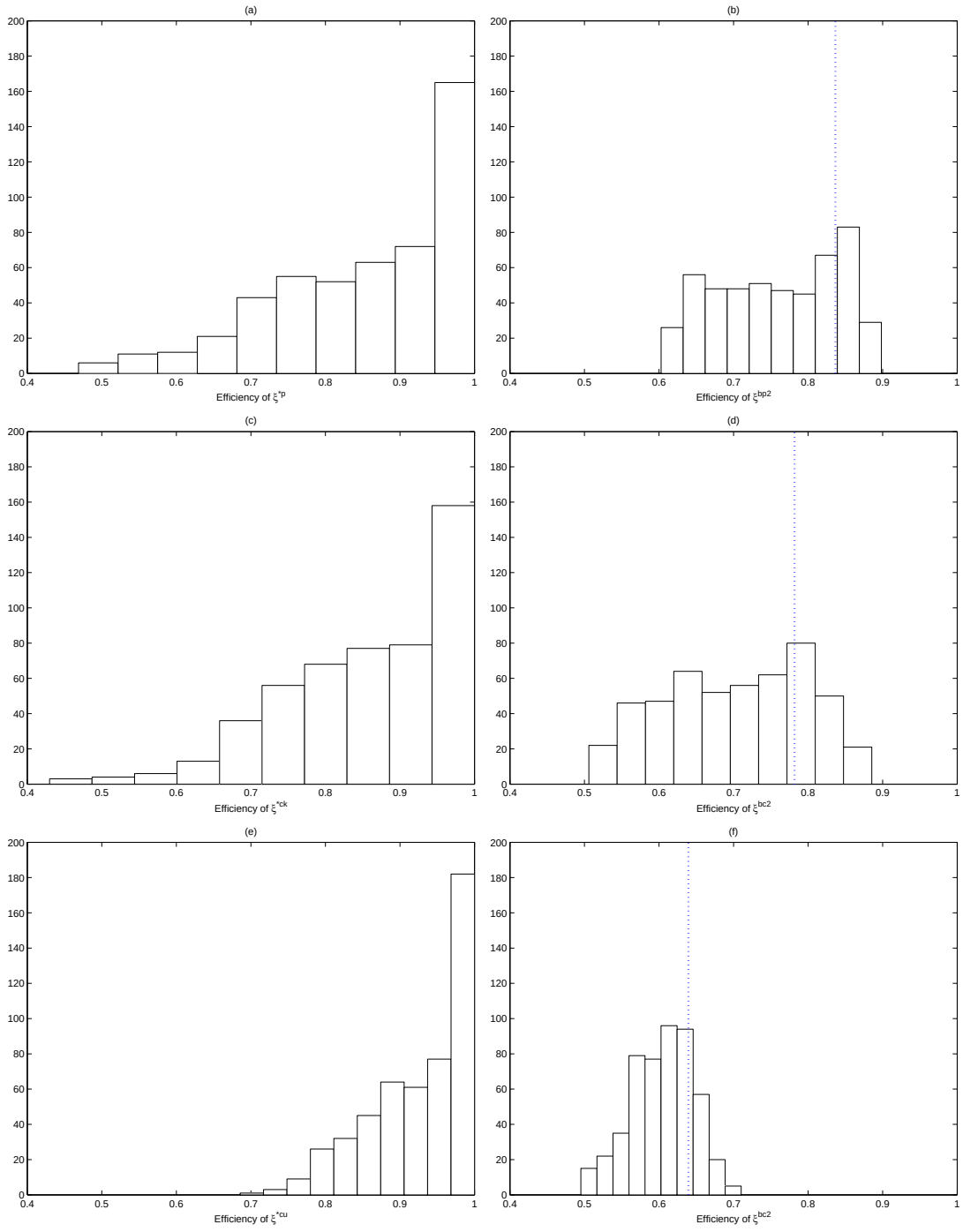

FIGURE 6.1: Histograms of efficiencies for the sensitivity analyses for optimal (left) and balanced (right) designs, for (a) & (b) parallel designs; (c) & (d) crossover designs where α is known; (e) & (f) crossover designs where α is unknown. The dotted lines on the histograms on the right hand side represent the efficiencies of the balanced designs for the original parameter values, $\boldsymbol{\theta} = (-1, 0.3, 0.25)'$.

receiving two doses, but if the loss of efficiency by including some subjects in a single period parallel group is not too great, this will allow for greater flexibility in the clinical trial. The more realistic model with α unknown is used for these designs.

We define a *union design* as a combination of each of the optimal designs for the parallel and crossover studies. The proportion of patients allocated to the parallel group is denoted by ρ. Given the optimal designs

$$\xi^{*p} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*p} & \dots & \boldsymbol{\xi}_{n_p}^{*p} \\ w_1^{*p} & \dots & w_{n_p}^{*p} \end{Bmatrix} \quad \text{and} \quad \xi^{*cu} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*cu} & \dots & \boldsymbol{\xi}_{n_{cu}}^{*cu} \\ w_1^{*cu} & \dots & w_{n_{cu}}^{*cu} \end{Bmatrix}$$

for the parallel and crossover models, respectively, we write the union design as

$$\xi_\rho^{\text{union}} = \left\{ \begin{array}{ccc:ccc} \boldsymbol{\xi}_1^{*p} & \dots & \boldsymbol{\xi}_{n_p}^{*p} & \boldsymbol{\xi}_1^{*cu} & \dots & \boldsymbol{\xi}_{n_{cu}}^{*cu} \\ \rho w_1^{*p} & \dots & \rho w_{n_p}^{*p} & (1-\rho) w_1^{*cu} & \dots & (1-\rho) w_{n_{cu}}^{*cu} \end{array} \right\}.$$

So $\xi_1^{\text{union}} = \xi^{*p}$ and $\xi_0^{\text{union}} = \xi^{*cu}$. This is similar to the idea of hybrid designs used in Chapters 3 and 5. For example, for equal allocation between the two groups ($\rho = 0.5$),

$$\xi_{0.5}^{\text{union}} = \left\{ \begin{array}{cc:ccc} 0.00 & 9.32 & (0.00, 0.00) & (0.00, 9.07) & (20.00, 0.00) \\ 0.25 & 0.25 & 0.020 & 0.286 & 0.194 \end{array} \right\}.$$

Since the model is the same as in section Section 6.4.2, the information matrix remains the same. It can be shown that $\boldsymbol{M}(\boldsymbol{\theta}, \xi_\rho^{\text{union}}) = \rho \boldsymbol{M}(\boldsymbol{\theta}, \xi^{*p}) + (1-\rho)\boldsymbol{M}(\boldsymbol{\theta}, \xi^{*cu})$. As mentioned previously, greater information will be gained by allocating a greater proportion of patients to the crossover arm of the trial. That is, the D-optimality criterion of ξ_ρ^{union} will be maximised for $\rho = 0$. Hence we define efficiencies of the union designs as

$$\text{Eff}(\xi_\rho^{\text{union}}) = \frac{|\boldsymbol{M}(\boldsymbol{\theta}, \xi_\rho^{\text{union}})|^{1/3}}{|\boldsymbol{M}(\boldsymbol{\theta}, \xi^{*cu})|^{1/3}}.$$

Efficiencies will obviously decrease as ρ approaches 1. However, parallel trials will generally be cheaper to conduct, as they require fewer doses and less of the patient's time. In this light, efficiencies must be adjusted for the reduction in cost resulting from allocating a patient to the parallel trial rather than the crossover trial. In order to do this, the information matrices must be adjusted for the reduction in cost. If C is the reduction in cost incurred by allocating a patient to a parallel design, the information gained by patients in the parallel group is inflated by the same amount:

$$\boldsymbol{M}_C(\boldsymbol{\theta}, \xi_\rho^{\text{union}}) = \rho C \boldsymbol{M}(\boldsymbol{\theta}, \xi^{*p}) + (1-\rho)\boldsymbol{M}(\boldsymbol{\theta}, \xi^{*cu}).$$

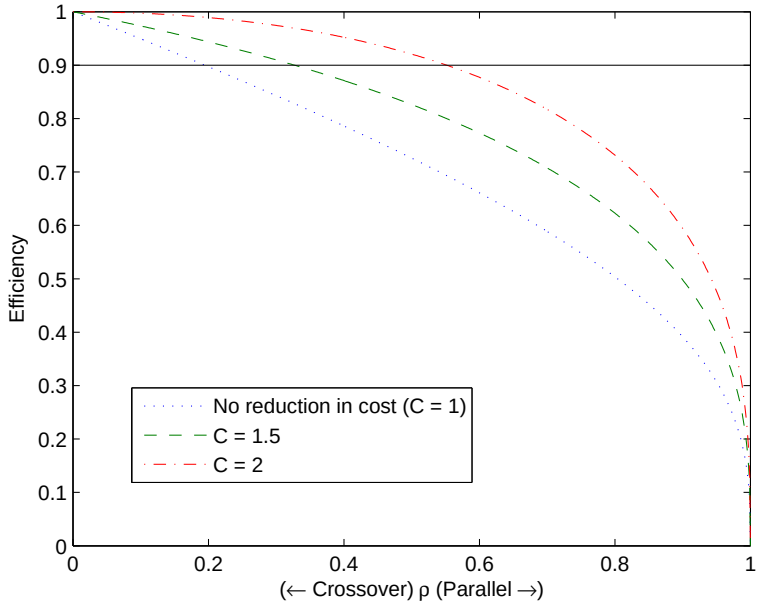

FIGURE 6.2: Efficiencies of union designs.

The efficiencies are adjusted accordingly:

$$\text{Eff}_C(\xi_\rho^{\text{union}}) = \frac{|\boldsymbol{M}_C(\boldsymbol{\theta}, \xi_\rho^{\text{union}})|^{1/3}}{|\boldsymbol{M}(\boldsymbol{\theta}, \xi^{*cu})|^{1/3}}.$$

These adjusted efficiencies are shown in Figure 6.2 for a range of ρ (from 0 to 1); and for $C = 1$ (no reduction in cost), $C = 1.5$ (a 50% reduction in cost), and $C = 2$ (the cost of the experiment is directly proportional to the number of doses used).

The experimenter may be interested in determining the maximum number of patients that can be allocated to the parallel group before the adjusted efficiency drops below a certain level, say 0.9. For the $C = 1$ case, no more than 19.4% of the patients can be to the parallel group before the efficiency of the design drops below 0.9. For $C = 1.5$ and $C = 2$, the maximum proportions of patients which can be placed in the parallel arm of the study increases to 32.6% and 55.1%, respectively. Depending on the magnitude of the extra cost involved with running a crossover trial, a good mixture of patients in the two treatment arms can be achieved without a great loss of efficiency.

6.7 Discussion

It has been shown that for a model with proportional carryover effects and binary outcomes, an optimal allocation of treatments is significantly more efficient than a balanced allocation, whether parallel designs, crossover designs or a combination of both in a union design are being considered. Further, it appears in light of the sensitivity analyses performed that the optimal designs are quite robust against parameter misspecification, and given that often experimenters have a reasonable idea of the prior distribution of the parameters, the results here are of practical value. They present experimenters with the potential of savings in terms of number of patients in an experiment without any loss in efficiency.

In addition to savings in terms of number of patients, the work here highlights that optimal designs may offer greater flexibility to the experimenter since with the union designs not all patients are required to receive both periods of treatments while these designs retain the advantages of the crossover structure. It is assumed in this setting that allocation to the parallel and crossover groups are randomised.

Although these findings are currently of theoretical interest only, at this stage, they do have significant potential for application in early clinical studies where significant PK data and some PD data may already be available and optimisation of first or second use in patients could be optimised further. The final design may well be the optimal design (perhaps even a union design) augmented with additional design points to ensure that the important clinical questions can be answered.

In the next chapter the effect of incorporating random coefficients in the models, to allow the model parameters to vary randomly between patients, is considered.

Chapter 7

Optimal designs for generalised linear mixed models

The GLMs in the previous two chapters have all been fixed effects models, i.e. all of the parameters are assumed to be fixed for all observations. In some circumstances we may wish to allow the parameters to vary between 'blocks' or 'clusters', and assume that observations within blocks (eg. subjects) are correlated. This is referred to as a mixed effects model. An example of a nonlinear mixed effects model is employed in Chapter 4, where the pharmacokinetic parameters such as clearance and volume are allowed to vary between patients, a much more realistic assumption than a single fixed clearance and volume for each patient.

If some parameters in a GLM are assumed to vary randomly between blocks, the model is known as a generalised linear mixed model, or GLMM. This chapter will focus on D-optimal design for GLMMs, in particular it is concerned with logistic regression models with random coefficients. The calculation of the information matrix presents a significant problem for GLMMs. The log-likelihood cannot be written down in closed form, hence either numerical methods (such as integration by quadrature) or approximations to the log-likelihood (such as the one used in Longford (1994)) must be used. Unfortunately, numerical methods can be very computationally intensive, even for very simple models and designs, and approximations can lead to large inaccuracies.

This chapter introduces an approximation to the information matrix which is not as

elegant as that of Longford (1994) (and hence has longer computation times), but is a more general approach and gives a significant increase in accuracy. Both approximations are then applied to the pharmacodynamic model used in the previous chapter, and the efficiencies of the resulting designs are compared.

7.1 Information matrix

Consider the logistic model with between-cluster variation for y_{ij}, the ith of n_j responses from the jth block (eg. subject), in Equation (2) of Longford (1994),

$$\text{logit}\{P(y_{ij}=1|\boldsymbol{\delta}_j)\} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}_j, \quad i=1,\ldots,n_j, \quad j=1,\ldots,N, \tag{7.1}$$

where the $\boldsymbol{z}_{ij}$ are subsets of the explanatory variables $\boldsymbol{x}_{ij}$ and elements of the $\boldsymbol{\delta}_j$ are independent standard normal random variables. Let $\boldsymbol{X}_j = (\boldsymbol{x}'_{1j},\ldots,\boldsymbol{x}'_{n_jj})'$ and similarly let $\boldsymbol{Z}_j = (\boldsymbol{z}'_{1j},\ldots,\boldsymbol{z}'_{n_j})'$. A distinction is made between the *fixed* parameters $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)'$ and the *random* parameters $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_q)'$ which are the unique elements of $\boldsymbol{\Sigma}$. The entire vector of parameters is then written as $\boldsymbol{\Psi} = (\boldsymbol{\beta}',\, \boldsymbol{\lambda}')'$.

The log-likelihood function for the jth block (with response vector $\boldsymbol{y}_j = (y_{1j},\ldots,y_{n_jj})'$) is written

$$\ell_j = \ell(\boldsymbol{\Psi},\boldsymbol{y}_j) = \log \int_{\mathbb{R}^q} \cdots \int P_j(\boldsymbol{\delta}_j)\Phi_q(\boldsymbol{\delta}_j)\mathrm{d}\boldsymbol{\delta}_j$$

where

$$\begin{aligned} P_j(\boldsymbol{\delta}_j) &= \prod_{i=1}^{n_j}\{\pi_{ij}(\boldsymbol{\delta}_j)\}^{y_{ij}}\{1-\pi_{ij}(\boldsymbol{\delta}_j)\}^{1-y_{ij}} \\ \pi_{ij}(\boldsymbol{\delta}_j) &= \text{logit}^{-1}\{\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}_j\} \end{aligned}$$

and Φ_q is the q-dimensional standard normal probability density function.

For an N-point exact design with response vector $\boldsymbol{y} = (\boldsymbol{y}'_1,\ldots,\boldsymbol{y}'_N)'$, the overall log-likelihood is

$$\ell = \ell(\boldsymbol{\Psi},\boldsymbol{y}) = \sum_{j=1}^{N}\ell(\boldsymbol{\Psi},\boldsymbol{y}_j),$$

and is given by

$$\ell = \ell(\boldsymbol{\Psi},\boldsymbol{y}) = \sum_{j=1}^{n} w_j\ell(\boldsymbol{\Psi},\boldsymbol{y}_j),$$

for an n-point approximate design, where w_j are the design weights.

Recall that the Fisher information matrix for the jth block, a function of the parameters $\boldsymbol{\Psi}$ and of the elementary design $\boldsymbol{\xi}_j$, is defined as

$$\begin{aligned} \boldsymbol{M}_F(\boldsymbol{\Psi}, \boldsymbol{\xi}_j) &= E\left(\frac{\partial \ell_j}{\partial \boldsymbol{\Psi}} \frac{\partial \ell_j}{\partial \boldsymbol{\Psi}'}\right) && (7.2) \\ &= -E\left(\frac{\partial^2 \ell_j}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'}\right) && (7.3) \\ &= -E\begin{bmatrix} \frac{\partial^2 \ell_j}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \frac{\partial^2 \ell_j}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}'} \\ \frac{\partial^2 \ell_j}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\beta}'} & \frac{\partial^2 \ell_j}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{A}_j & \boldsymbol{C}_j \\ \boldsymbol{C}_j' & \boldsymbol{B}_j \end{bmatrix}, && (7.4) \end{aligned}$$

7.1.1 Numerical methods

For an exact design, the expectation in Equation (7.3) may be calculated by the following sum:

$$\mathrm{E}\left[\frac{\partial^2 \ell(\boldsymbol{\Psi}, y)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'}\right] = \sum_{y_{ij}} \frac{\partial^2 \ell(\boldsymbol{\Psi}, y_{ij})}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'} \ell(\boldsymbol{\Psi}, y_{ij})$$

where the sum is over the $2^{\sum n_j}$ possible outcomes for $\boldsymbol{y}_j$, and the derivatives are evaluated numerically, using finite difference methods (Press *et al.*, 2002). The log-likelihood may be calculated by either of two ways: by adaptive Lobatto quadrature (Gander and Gautschi, 2000) via the `quadl` function in MATLAB when $q = 1$, in which case the information matrix (assumed to be an accurate representation of the 'true' information matrix) is denoted by $\boldsymbol{M}_N(\boldsymbol{\Psi}, \xi)$; or by the same approximation to the log-likelihood used in the next two methods, in which case the information matrix is denoted by $\boldsymbol{M}_{Na}(\boldsymbol{\Psi}, \xi)$. It is assumed that $\boldsymbol{M}_N(\boldsymbol{\Psi}, \xi)$ is an accurate representation of the 'true' information matrix, as it contains no approximations, and any numerical error associated with the differentiation and integration is assumed to be negligible.

These numerical methods are only applicable to exact designs, as they require summation over possible outcomes for $\boldsymbol{y}_j$ at each support point. This is not defined for an approximate design.

7.1.2 Longford approximation

It can be shown that the log-likelihood can be approximated by the following second-order Taylor expansion:

$$\begin{aligned}\ell &= \sum_{j=1}^{N} \ell_j \\ &\approx \sum_{j=1}^{N} \left[\log\{P_j(\mathbf{0})\} + \frac{1}{2}\boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j - \frac{1}{2}\log|\boldsymbol{G}_j| \right],\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{G}_j &= \boldsymbol{I}_q + \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}, \\ \boldsymbol{H}_j &= \boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j, \\ \boldsymbol{W}_j &= \mathrm{diag}\{\boldsymbol{w}_j\}, \\ \boldsymbol{w}_j &= (w_{1j}, \ldots, w_{n_j j})', \quad \text{and} \\ w_{ij} &= \pi_{ij}(0)(1-\pi_{ij}(0)).\end{aligned}$$

Using this approximate log-likelihood, Longford (1994) finds the expectation of its derivatives to give an approximation to $\boldsymbol{A}_j$, $\boldsymbol{B}_j$ and $\boldsymbol{C}_j$ in Equation (7.4) by ignoring the dependence of $w_{ij} = \pi_{ij}(1-\pi_{ij})$ on $\boldsymbol{\beta}$:

$$\boldsymbol{A}_j = \boldsymbol{X}_j'\boldsymbol{V}_j^{-1}\boldsymbol{X}_j,$$

where $\boldsymbol{V}_j = \boldsymbol{W}_j^{-1} + \boldsymbol{Z}_j\boldsymbol{\Sigma}\boldsymbol{Z}_j'$; the (u, v) element of $\boldsymbol{B}_j$ is

$$\boldsymbol{B}_j^{(u,v)} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_j^{-1}\frac{\partial \boldsymbol{V}_j}{\partial \sigma_u^2}\boldsymbol{V}_j^{-1}\frac{\partial \boldsymbol{V}_j}{\partial \sigma_v^2} \right);$$

and $\boldsymbol{C}_j = \mathbf{0}$. The information matrix calculated by this method is denoted by $\boldsymbol{M}_L(\boldsymbol{\Psi}, \xi)$.

Some investigation (presented later in this chapter) using the numerical methods outlined in Section 7.1.1 have shown that such an assumption may lead to significant error. In the next section, a more accurate approximation to Equation (7.4) is given by acknowledging the dependence of w_{ij} on $\boldsymbol{\beta}$.

7.1.3 Approximation acknowledging dependence of w_{ij} on $\boldsymbol{\beta}$

The approximation to the information matrix presented in this section is denoted by $\boldsymbol{M}_W(\boldsymbol{\Psi}, \xi)$, where it is hoped that the W in the subscript will remind readers of the dependence of w_{ij}

(elements of the $\boldsymbol{W}_j$ matrix) on $\boldsymbol{\beta}$. It is hoped that acknowledging this dependence will produce a more accurate approximation to the information matrix. In particular, it will account for interaction of the fixed and random effects by finding a nonzero matrix $\boldsymbol{C}_j$ in Equation (7.4).

For any given parameters ψ_m and ψ_n,

$$\frac{\partial^2 \ell_j}{\partial\psi_m\partial\psi_n} \approx \underbrace{\frac{\partial^2 \log\{P_j(\mathbf{0})\}}{\partial\psi_m\partial\psi_n}}_{\text{Term 1}} + \frac{1}{2}\underbrace{\frac{\partial^2 \boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j}{\partial\psi_m\partial\psi_n}}_{\text{Term 2}} - \frac{1}{2}\underbrace{\frac{\partial^2 \log|\boldsymbol{G}_j|}{\partial\psi_m\partial\psi_n}}_{\text{Term 3}}$$

Each of these three terms in turn are considered in turn.

Term 1

Firstly, $P_j(\mathbf{0}) = \prod_{i=1}^{n_j} \pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}}$ does not involve any of the random parameters, so for any $m = 1, \ldots, q$,

$$\frac{\partial \log\{P_j(\mathbf{0})\}}{\partial\lambda_m} = 0.$$

So clearly

$$\begin{aligned}
\frac{\partial^2 \log\{P_j(\mathbf{0})\}}{\partial\lambda_m\partial\lambda_n} &= 0, \quad m,\, n = 1, \ldots, q, \quad \text{and} \\
\frac{\partial^2 \log\{P_j(\mathbf{0})\}}{\partial\lambda_m\partial\beta_n} &= 0, \quad m = 1, \ldots, q,\, n = 1, \ldots, p.
\end{aligned}$$

Now consider one of the fixed parameters, β_m:

$$\begin{aligned}
\frac{\partial \log\{P_j(\mathbf{0})\}}{\partial\beta_m} &= \frac{\partial}{\partial\beta_m}\left[\log\prod_{i=1}^{n_j}\pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}}\right] \\
&= \sum_i \frac{\partial}{\partial\beta_m}\left[y_{ij}\log\pi_{ij} + (1-y_{ij})\log(1-\pi_{ij})\right] \\
&= \sum_i \frac{\partial}{\partial\beta_m}\left[y_{ij}\log\left\{\frac{\pi_{ij}}{1-\pi_{ij}}\right\} + \log(1-\pi_{ij})\right] \\
&= \sum_i \left[y_{ij}\frac{\partial \boldsymbol{x}_{ij}\boldsymbol{\beta}}{\partial\beta_m} + \frac{\partial\log(1-\pi_{ij})}{\partial\beta_m}\right] \\
&= \sum_i \left[y_{ij}x_{ijm} - \left(\frac{1}{1-\pi_{ij}}\right)\frac{\partial\pi_{ij}}{\partial\beta_m}\right]
\end{aligned}$$

Now,

$$\begin{aligned}\frac{\partial \pi_{ij}}{\partial \beta_m} &= \frac{\partial}{\partial \beta_m}\left[\frac{1}{1+\exp\{-\boldsymbol{x}_{ij}\boldsymbol{\beta}\}}\right] \\ &= x_{ijm}\left[\frac{\exp\{-\boldsymbol{x}_{ij}\boldsymbol{\beta}\}}{(1+\exp\{-\boldsymbol{x}_{ij}\boldsymbol{\beta}\})^2}\right] \\ &= x_{ijm}\pi_{ij}(1-\pi_{ij}) \\ &= x_{ijm}w_{ij}\end{aligned}$$

So

$$\begin{aligned}\frac{\partial \log\{P_j(\mathbf{0})\}}{\partial \beta_m} &= \sum_i \left[y_{ij}x_{ijm} - x_{ijm}\pi_{ij}\right] \\ &= \sum_i x_{ijm}e_{ij}w_{ij}\end{aligned}$$

Taking the second derivative,

$$\begin{aligned}\frac{\partial^2 \log\{P_j(\mathbf{0})\}}{\partial \beta_m \partial \beta_n} &= \frac{\partial}{\partial \beta_n}\sum_i \left[y_{ij}x_{ijm} - x_{ijm}\pi_{ij}\right] \\ &= \sum_i \left[-x_{ijm}\frac{\partial \pi_{ij}}{\partial \beta_n}\right] \\ &= -\sum_i x_{ijm}x_{ijn}w_{ij} \\ &= -\boldsymbol{X}'_{j(m)}\boldsymbol{W}_j\boldsymbol{X}_{j(n)}\end{aligned}$$

where $\boldsymbol{X}_{j(m)}$ is the mth column of $\boldsymbol{X}_j$.

Term 2

For any parameter ψ_m,

$$\begin{aligned}\frac{\partial \boldsymbol{e}'_j\boldsymbol{H}_j\boldsymbol{e}_j}{\partial \psi_m} &= \boldsymbol{e}'_j\left(\frac{\partial \boldsymbol{H}_j\boldsymbol{e}_j}{\partial \psi_m}\right) + \left(\frac{\partial \boldsymbol{e}'_j}{\partial \psi_m}\right)\boldsymbol{H}_j\boldsymbol{e}_j \qquad (7.5) \\ &= \boldsymbol{e}'_j\left(\boldsymbol{H}_j\frac{\partial \boldsymbol{e}_j}{\partial \psi_m} + \frac{\partial \boldsymbol{H}_j}{\partial \psi_m}\boldsymbol{e}_j\right) + \frac{\partial \boldsymbol{e}'_j}{\partial \psi_m}\boldsymbol{H}_j\boldsymbol{e}_j \\ &= \boldsymbol{e}'_j\boldsymbol{H}_j\frac{\partial \boldsymbol{e}_j}{\partial \psi_m} + \boldsymbol{e}'_j\frac{\partial \boldsymbol{H}_j}{\partial \psi_m}\boldsymbol{e}_j + \frac{\partial \boldsymbol{e}'_j}{\partial \psi_m}\boldsymbol{H}_j\boldsymbol{e}_j \\ &= 2\boldsymbol{e}'_j\boldsymbol{H}_j\frac{\partial \boldsymbol{e}_j}{\partial \psi_m} + \boldsymbol{e}'_j\frac{\partial \boldsymbol{H}_j}{\partial \psi_m}\boldsymbol{e}_j \qquad (7.6)\end{aligned}$$

The last step is true because

$$\begin{aligned}
\boldsymbol{e}_j' \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} &= \left(\boldsymbol{e}_j' \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \right)' \qquad \text{(scalar)} \\
&= \left(\frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \right)' \boldsymbol{H}_j' (\boldsymbol{e}_j')' \\
&= \frac{\partial \boldsymbol{e}_j'}{\partial \psi_m} \boldsymbol{H}_j \boldsymbol{e}_j \qquad (\boldsymbol{H}_j \text{ symmetric})
\end{aligned}$$

The second derivative:

$$\frac{\partial^2 \boldsymbol{e}_j' \boldsymbol{H}_j \boldsymbol{e}_j}{\partial \psi_m \partial \psi_n} = \frac{\partial}{\partial \psi_n} \left[2 \boldsymbol{e}_j' \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} + \boldsymbol{e}_j' \frac{\partial \boldsymbol{H}_j}{\partial \psi_m} \boldsymbol{e}_j \right]$$

Now,

$$\begin{aligned}
\frac{\partial}{\partial \psi_n} \left[\boldsymbol{e}_j' \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \right] &= \boldsymbol{e}_j' \frac{\partial}{\partial \psi_n} \left[\boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \right] + \frac{\partial \boldsymbol{e}_j'}{\partial \psi_n} \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \\
&= \boldsymbol{e}_j' \left[\boldsymbol{H}_j \frac{\partial^2 \boldsymbol{e}_j}{\partial \psi_m \partial \psi_n} + \frac{\partial \boldsymbol{H}_j}{\partial \psi_n} \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \right] + \frac{\partial \boldsymbol{e}_j'}{\partial \psi_n} \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \\
&= \boldsymbol{e}_j' \boldsymbol{H}_j \frac{\partial^2 \boldsymbol{e}_j}{\partial \psi_m \partial \psi_n} + \boldsymbol{e}_j' \frac{\partial \boldsymbol{H}_j}{\partial \psi_n} \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} + \frac{\partial \boldsymbol{e}_j'}{\partial \psi_n} \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m}
\end{aligned}$$

And, similarly to Equations (7.5)–(7.6),

$$\frac{\partial}{\partial \psi_n} \left[\boldsymbol{e}_j' \frac{\partial \boldsymbol{H}_j}{\partial \psi_m} \boldsymbol{e}_j \right] = 2 \boldsymbol{e}_j' \frac{\partial \boldsymbol{H}_j}{\partial \psi_m} \frac{\partial \boldsymbol{e}_j}{\partial \psi_n} + \boldsymbol{e}_j' \frac{\partial^2 \boldsymbol{H}_j}{\partial \psi_m \partial \psi_n} \boldsymbol{e}_j$$

So then

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{e}_j' \boldsymbol{H}_j \boldsymbol{e}_j}{\partial \psi_m \partial \psi_n} &= 2 \left[\boldsymbol{e}_j' \boldsymbol{H}_j \frac{\partial^2 \boldsymbol{e}_j}{\partial \psi_m \partial \psi_n} + \boldsymbol{e}_j' \frac{\partial \boldsymbol{H}_j}{\partial \psi_n} \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} + \frac{\partial \boldsymbol{e}_j'}{\partial \psi_n} \boldsymbol{H}_j \frac{\partial \boldsymbol{e}_j}{\partial \psi_m} \right] + \\
&\quad + 2 \boldsymbol{e}_j' \frac{\partial \boldsymbol{H}_j}{\partial \psi_m} \frac{\partial \boldsymbol{e}_j}{\partial \psi_n} + \boldsymbol{e}_j' \frac{\partial^2 \boldsymbol{H}_j}{\partial \psi_m \partial \psi_n} \boldsymbol{e}_j
\end{aligned} \tag{7.7}$$

Fixed parameters only

$$\frac{\partial \boldsymbol{e}_j}{\partial \beta_m} = \frac{\partial}{\partial \beta_m} \left[e_{1j}, \ldots, e_{n_j j} \right]'$$

Now,

$$\begin{aligned}
\frac{\partial e_{ij}}{\partial \beta_m} &= \frac{\partial}{\partial \beta_m} \left[\frac{(y_{ij} - \pi_{ij})}{w_{ij}} \right] \\
&= \frac{w_{ij} \left(-\frac{\partial \pi_{ij}}{\partial \beta_m} \right) - (y_{ij} - \pi_{ij}) \left(\frac{\partial w_{ij}}{\partial \beta_m} \right)}{w_{ij}^2}
\end{aligned}$$

We have

$$\frac{\partial w_{ij}}{\partial \beta_m} = \frac{\partial}{\partial \beta_m}\left[\pi_{ij}(1-\pi_{ij})\right] = \frac{\partial \pi_{ij}}{\partial \beta_m} - 2\pi_{ij}\frac{\partial \pi_{ij}}{\partial \beta_m} = x_{ijm}w_{ij}(1-2\pi_{ij}) \tag{7.8}$$

So

$$\begin{aligned}
\frac{\partial e_{ij}}{\partial \beta_m} &= \frac{-x_{ijm}w_{ij}^2 - (y_{ij}-\pi_{ij})x_{ijm}w_{ij}(1-2\pi_{ij})}{w_{ij}^2} \\
&= -x_{ijm}[1+e_{ij}(1-2\pi_{ij})] \\
\frac{\partial \boldsymbol{e}_j}{\partial \beta_m} &= -\boldsymbol{X}_{j(m)} - \boldsymbol{T}_{1jm}\boldsymbol{e}_j
\end{aligned}$$

where $\boldsymbol{T}_{1jm} = \text{diag}(x_{1jm}(1-2\pi_{1j}), \ldots, x_{n_j jm}(1-2\pi_{n_j j}))$. So it follows that

$$\begin{aligned}
\frac{\partial^2 e_{ij}}{\partial \beta_m \partial \beta_n} &= \frac{\partial}{\partial \beta_n}\left[-x_{ijm}(1+e_{ij}(1-2\pi_{ij}))\right] \\
&= -x_{ijm}\left[e_{ij}\left(-2\frac{\partial \pi_{ij}}{\partial \beta_n}\right) + \frac{\partial e_{ij}}{\partial \beta_n}(1-2\pi_{ij})\right] \\
&= x_{ijm}\left[e_{ij}(2x_{ijn}w_{ij}) + x_{ijn}[1+e_{ij}(1-2\pi_{ij})](1-2\pi_{ij})\right] \\
&= x_{ijm}\left[2x_{ijn}w_{ij}e_{ij} + x_{ijn}(1+e_{ij}-2e_{ij}\pi_{ij})(1-2\pi_{ij})\right] \\
&= x_{ijm}[2x_{ijn}w_{ij}e_{ij} + \\
&\quad + x_{ijn}(1-2\pi_{ij}+e_{ij}-2e_{ij}\pi_{ij}-2e_{ij}\pi_{ij}+4e_{ij}\pi_{ij}^2)] \\
&= x_{ijm}[e_{ij}(2x_{ijn}w_{ij} + x_{ijn}(1-4\pi_{ij}+4\pi_{ij}^2)) + \\
&\quad + x_{ijn}(1-2\pi_{ij})] \\
&= e_{ij}x_{ijm}x_{ijn}(2w_{ij} + (1-2\pi_{ij})^2) + \\
&\quad + x_{ijm}x_{ijn}(1-2\pi_{ij}) \\
\frac{\partial^2 \boldsymbol{e}_j}{\partial \beta_m \partial \beta_n} &= \boldsymbol{T}_{2jmn} + \boldsymbol{T}_{3jmn}\boldsymbol{e}_j
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{T}_{2jmn} &= (x_{1jm}x_{1jn}(1-2\pi_{1j}), \ldots, x_{n_j jm}x_{n_j jn}(1-2\pi_{n_j j}))' \quad \text{and} \\
\boldsymbol{T}_{3jmn} &= \text{diag}(x_{1jm}x_{1jn}(2w_{1j} + (1-2\pi_{1j})^2), \ldots, \\
&\qquad x_{n_j jm}x_{n_j jn}(2w_{n_j j} + (1-2\pi_{n_j j})^2)).
\end{aligned}$$

Considering each term of Equation (7.7),

$$\begin{aligned}
\boldsymbol{e}_j'\boldsymbol{H}_j\frac{\partial^2\boldsymbol{e}_j}{\partial\beta_m\partial\beta_n} &= \boldsymbol{e}_j'\boldsymbol{H}_j\left(\boldsymbol{T}_{2jmn}+\boldsymbol{T}_{3jmn}\boldsymbol{e}_j\right)\\
&= \boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{T}_{2jmn}+\boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{T}_{3jmn}\boldsymbol{e}_j\\
\boldsymbol{e}_j'\frac{\partial\boldsymbol{H}_j}{\partial\beta_n}\frac{\partial\boldsymbol{e}_j}{\partial\beta_m} &= \boldsymbol{e}_j'\frac{\partial\boldsymbol{H}_j}{\partial\beta_n}\left(-\boldsymbol{X}_{j(m)}-\boldsymbol{T}_{1jm}\boldsymbol{e}_j\right)\\
&= -\boldsymbol{e}_j'\frac{\partial\boldsymbol{H}_j}{\partial\beta_n}\boldsymbol{X}_{j(m)}-\boldsymbol{e}_j'\frac{\partial\boldsymbol{H}_j}{\partial\beta_n}\boldsymbol{T}_{1jm}\boldsymbol{e}_j\\
\frac{\partial\boldsymbol{e}_j'}{\partial\beta_n}\boldsymbol{H}_j\frac{\partial\boldsymbol{e}_j}{\partial\beta_m} &= \left(-\boldsymbol{X}_{j(n)}-\boldsymbol{T}_{1jn}\boldsymbol{e}_j\right)'\boldsymbol{H}_j\left(-\boldsymbol{X}_{j(m)}-\boldsymbol{T}_{1jm}\boldsymbol{e}_j\right)\\
&= \boldsymbol{X}_{j(n)}'\boldsymbol{H}_j\boldsymbol{X}_{j(m)}+\boldsymbol{e}_j'\boldsymbol{T}_{1jn}'\boldsymbol{H}_j\boldsymbol{T}_{1jm}\boldsymbol{e}_j+\\
&\quad+\boldsymbol{X}_{j(n)}'\boldsymbol{H}_j\boldsymbol{T}_{1jm}\boldsymbol{e}_j+\boldsymbol{e}_j'\boldsymbol{T}_{1jn}'\boldsymbol{H}_j\boldsymbol{X}_{j(m)}
\end{aligned}$$

So Equation (7.7) becomes

$$\begin{aligned}
\frac{\partial^2\boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j}{\partial\beta_m\partial\beta_n} &= 2\boldsymbol{e}_j'\Bigg[\boldsymbol{H}_j\boldsymbol{T}_{2jmn}-\frac{\partial\boldsymbol{H}_j}{\partial\beta_n}\boldsymbol{X}_{j(m)}+\boldsymbol{T}_{1jm}'\boldsymbol{H}_j'\boldsymbol{X}_{j(n)}+\\
&\qquad+\boldsymbol{T}_{1jn}'\boldsymbol{H}_j\boldsymbol{X}_{j(m)}-\frac{\partial\boldsymbol{H}_j}{\partial\beta_m}\boldsymbol{X}_{j(n)}\Bigg]+\\
&\quad+\boldsymbol{e}_j'\Bigg[2\boldsymbol{H}_j\boldsymbol{T}_{3jmn}-2\frac{\partial\boldsymbol{H}_j}{\partial\beta_n}\boldsymbol{T}_{1jm}+2\boldsymbol{T}_{1jn}'\boldsymbol{H}_j\boldsymbol{T}_{1jm}-\\
&\qquad-2\frac{\partial\boldsymbol{H}_j}{\partial\beta_m}\boldsymbol{T}_{1jn}+\frac{\partial^2\boldsymbol{H}_j}{\partial\beta_m\partial\beta_n}\Bigg]\boldsymbol{e}_j+\\
&\quad+2\boldsymbol{X}_{j(n)}'\boldsymbol{H}_j\boldsymbol{X}_{j(m)}
\end{aligned}$$

Next,

$$\begin{aligned}
\frac{\partial\boldsymbol{H}_j}{\partial\beta_m} &= \boldsymbol{W}_j\left(\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}+\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_m}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j\right)+\\
&\quad+\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j\\
&= \boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}+\boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_m}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j+\\
&\quad+\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j
\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \boldsymbol{H}_j}{\partial\beta_m\partial\beta_n} &= \boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial^2\boldsymbol{W}_j}{\partial\beta_m\partial\beta_n} + \boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_n}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m} + \\ &+ \frac{\partial\boldsymbol{W}_j}{\partial\beta_n}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m} + \boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_m}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_n} + \\ &+ \boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\frac{\partial^2\boldsymbol{G}_j^{-1}}{\partial\beta_m\partial\beta_n}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j + \frac{\partial\boldsymbol{W}_j}{\partial\beta_n}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_m}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j + \\ &+ \frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_n} + \frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_n}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j + \\ &+ \frac{\partial^2\boldsymbol{W}_j}{\partial\beta_m\partial\beta_n}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j\end{aligned}$$

Random parameters only

Since $\boldsymbol{e}_j$ does not depend on any random parameters, Equation (7.7) reduces to simply

$$\frac{\partial^2\boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j}{\partial\lambda_m\partial\lambda_n} = \boldsymbol{e}_j'\frac{\partial^2\boldsymbol{H}_j}{\partial\lambda_m\partial\lambda_n}\boldsymbol{e}_j$$

$$\begin{aligned}\frac{\partial\boldsymbol{H}_j}{\partial\lambda_m} &= \boldsymbol{W}_j\boldsymbol{Z}_j\left[\boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m} + \frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_m}\boldsymbol{\Sigma}^{1/2}\right) + \frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\right]\boldsymbol{Z}_j'\boldsymbol{W}_j \\ &= \boldsymbol{W}_j\boldsymbol{Z}_j\left[\boldsymbol{\Sigma}^{1/2}\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m} + \boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_m}\boldsymbol{\Sigma}^{1/2} + \frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\right]\boldsymbol{Z}_j'\boldsymbol{W}_j\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\boldsymbol{H}_j}{\partial\lambda_m\partial\lambda_n} &= \boldsymbol{W}_j\boldsymbol{Z}_j\left[\boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{G}_j^{-1}\frac{\partial^2\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m\partial\lambda_n} + \frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_n}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}\right) + \frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m} + \right. \\ &+ \boldsymbol{\Sigma}^{1/2}\left(\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_m}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n} + \frac{\partial^2\boldsymbol{G}_j^{-1}}{\partial\lambda_m\partial\lambda_n}\boldsymbol{\Sigma}^{1/2}\right) + \frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_n}\boldsymbol{\Sigma}^{1/2} + \\ &\left.+ \frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}\left(\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n} + \frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_n}\boldsymbol{\Sigma}^{1/2}\right) + \frac{\partial^2\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m\partial\lambda_n}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\right]\boldsymbol{Z}_j'\boldsymbol{W}_j\end{aligned}$$

One fixed parameter, one random parameter

Equation (7.7) simplifies to

$$\begin{aligned}\frac{\partial^2\boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j}{\partial\beta_m\partial\lambda_n} &= \boldsymbol{e}_j'\frac{\partial^2\boldsymbol{H}_j}{\partial\beta_m\partial\lambda_n}\boldsymbol{e}_j + 2\boldsymbol{e}_j'\frac{\partial\boldsymbol{H}_j}{\partial\lambda_n}\frac{\partial\boldsymbol{e}_j}{\partial\beta_m} \\ &= \boldsymbol{e}_j'\frac{\partial^2\boldsymbol{H}_j}{\partial\beta_m\partial\lambda_n}\boldsymbol{e}_j + 2\boldsymbol{e}_j'\frac{\partial\boldsymbol{H}_j}{\partial\lambda_n}\left(-\boldsymbol{X}_{j(m)} - \boldsymbol{T}_{1jm}\boldsymbol{e}_j\right) \\ &= -2\boldsymbol{e}_j'\frac{\partial\boldsymbol{H}_j}{\partial\lambda_n}\boldsymbol{X}_{j(m)} + \boldsymbol{e}_j'\left(\frac{\partial^2\boldsymbol{H}_j}{\partial\beta_m\partial\lambda_n} - 2\frac{\partial\boldsymbol{H}_j}{\partial\lambda_n}\boldsymbol{T}_{1jm}\right)\boldsymbol{e}_j\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{H}_j}{\partial\beta_m\partial\lambda_n} &= \boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}+\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_n}\boldsymbol{\Sigma}^{1/2}\right)\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}+\\
&+\boldsymbol{W}_j\boldsymbol{Z}_j\left(\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}+\boldsymbol{\Sigma}^{1/2}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_m}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\boldsymbol{Z}_j'\boldsymbol{W}_j\right)+\\
&+\boldsymbol{W}_j\boldsymbol{Z}_j\left(\boldsymbol{\Sigma}^{1/2}\frac{\partial^2\boldsymbol{G}_j^{-1}}{\partial\beta_m\partial\lambda_n}+\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\beta_m}\right)\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j+\\
&+\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}+\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\lambda_n}\boldsymbol{\Sigma}^{1/2}\right)\boldsymbol{Z}_j'\boldsymbol{W}_j+\\
&+\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\boldsymbol{G}_j^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j
\end{aligned}$$

Derivatives of $\boldsymbol{G}_j^{-1}$

For general parameters ψ_m and ψ_n,

$$\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\psi_m}=-\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{G}_j}{\partial\psi_m}\boldsymbol{G}_j^{-1}$$

$$\begin{aligned}
\frac{\partial^2\boldsymbol{G}_j^{-1}}{\partial\psi_m\partial\psi_n} &= -\left[\boldsymbol{G}_j^{-1}\left(\frac{\partial\boldsymbol{G}_j}{\partial\psi_m}\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\psi_n}+\frac{\partial^2\boldsymbol{G}_j}{\partial\psi_m\partial\psi_n}\boldsymbol{G}_j^{-1}\right)+\frac{\partial\boldsymbol{G}_j^{-1}}{\partial\psi_n}\frac{\partial\boldsymbol{G}_j}{\partial\psi_m}\boldsymbol{G}_j^{-1}\right]\\
&= -\left[\boldsymbol{G}_j^{-1}\left(-\frac{\partial\boldsymbol{G}_j}{\partial\psi_m}\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{G}_j}{\partial\psi_n}\boldsymbol{G}_j^{-1}+\frac{\partial^2\boldsymbol{G}_j}{\partial\psi_m\partial\psi_n}\boldsymbol{G}_j^{-1}\right)-\right.\\
&\quad\left.-\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{G}_j}{\partial\psi_n}\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{G}_j}{\partial\psi_m}\boldsymbol{G}_j^{-1}\right]\\
&= \boldsymbol{G}_j^{-1}\left(\frac{\partial\boldsymbol{G}_j}{\partial\psi_m}\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{G}_j}{\partial\psi_n}-\frac{\partial^2\boldsymbol{G}_j}{\partial\psi_m\partial\psi_n}+\frac{\partial\boldsymbol{G}_j}{\partial\psi_n}\boldsymbol{G}_j^{-1}\frac{\partial\boldsymbol{G}_j}{\partial\psi_m}\right)\boldsymbol{G}_j^{-1}
\end{aligned}$$

For fixed parameters β_m and β_n, and random parameters λ_m and λ_n,

$$\frac{\partial\boldsymbol{G}_j}{\partial\beta_m}=\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2} \tag{7.9}$$

$$\begin{aligned}
\frac{\partial^2\boldsymbol{G}_j}{\partial\beta_m\partial\beta_n} &= \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial^2\boldsymbol{W}_j}{\partial\beta_m\partial\beta_n}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\\
\frac{\partial^2\boldsymbol{G}_j}{\partial\beta_m\partial\lambda_n} &= \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}+\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\boldsymbol{Z}_j'\frac{\partial\boldsymbol{W}_j}{\partial\beta_m}\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\\
\frac{\partial\boldsymbol{G}_j}{\partial\lambda_m} &= \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j\boldsymbol{Z}_j\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}+\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}\boldsymbol{Z}_j'\boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}\\
\frac{\partial^2\boldsymbol{G}_j}{\partial\lambda_m\partial\lambda_n} &= \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_j'\boldsymbol{W}_j\boldsymbol{Z}_j\frac{\partial^2\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m\partial\lambda_n}+\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}\boldsymbol{Z}_j'\boldsymbol{W}_j\boldsymbol{Z}_j\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}+\\
&\quad+\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m}\boldsymbol{Z}_j'\boldsymbol{W}_j\boldsymbol{Z}_j\frac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_n}+\frac{\partial^2\boldsymbol{\Sigma}^{1/2}}{\partial\lambda_m\partial\lambda_n}\boldsymbol{Z}_j'\boldsymbol{W}_j\boldsymbol{Z}_j\boldsymbol{\Sigma}^{1/2}
\end{aligned} \tag{7.10}$$

Derivatives of $\boldsymbol{\Sigma}^{1/2}$

It is assumed that $\boldsymbol{\Sigma} = \text{diag}(\lambda_1, \ldots, \lambda_q)$, so $\boldsymbol{\Sigma}^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_q^{1/2})$. So we have

$$\left[\frac{\partial \boldsymbol{\Sigma}^{1/2}}{\partial \lambda_m}\right]_{(i,j)} = \begin{cases} \frac{1}{2}\lambda_m^{-1/2} & \text{if} \quad i = j = m \\ \\ 0 & \text{otherwise} \end{cases}.$$

For the second derivatives,

$$\left[\frac{\partial^2 \boldsymbol{\Sigma}^{1/2}}{\partial \lambda_m^2}\right]_{(i,j)} = \begin{cases} -\frac{1}{4}\lambda_m^{-3/2} & \text{if} \quad i = j = m \\ \\ 0 & \text{otherwise} \end{cases}$$

and, for $m \neq n$,

$$\frac{\partial^2 \boldsymbol{\Sigma}^{1/2}}{\partial \lambda_m \partial \lambda_n} = \mathbf{0}_{q \times q}.$$

Derivatives of $\boldsymbol{W}_j$

We already have from Equation (7.8) that

$$\frac{\partial w_{ij}}{\partial \beta_m} = x_{ijm} w_{ij}(1 - 2\pi_{ij}),$$

so

$$\begin{aligned} \frac{\partial^2 w_{ij}}{\partial \beta_m \partial \beta_n} &= x_{ijm}\left[w_{ij}\left(-2\frac{\partial \pi_{ij}}{\partial \beta_n}\right) + \frac{\partial w_{ij}}{\partial \beta_n}(1 - 2\pi_{ij})\right] \\ &= x_{ijm}\left[w_{ij}\left(-2x_{ijn}w_{ij}\right) + x_{ijn}w_{ij}\left(1 - 2\pi_{ij}\right)^2\right] \\ &= x_{ijm}w_{ij}\left[-2x_{ijn}w_{ij} + x_{ijn}\left(1 - 2\pi_{ij}\right)^2\right] \\ &= x_{ijm}w_{ij}\left[-x_{ijn}\left(2\pi_{ij} - 2\pi_{ij}^2 - 1 + 4\pi_{ij} - 4\pi_{ij}^2\right)\right] \\ &= x_{ijm}w_{ij}\left[x_{ijn}\left(1 - 6\pi_{ij} + 6\pi_{ij}^2\right)\right] \\ &= x_{ijm}x_{ijn}w_{ij}\left(1 - 6w_{ij}\right) \end{aligned}$$

Term 3

For general parameters ψ_m and ψ_n,

$$\frac{\partial \log|\boldsymbol{G}_j|}{\partial \psi_m} = \text{tr}\left(\boldsymbol{G}_j^{-1}\frac{\partial \boldsymbol{G}_j}{\partial \psi_m}\right)$$

$$\begin{aligned}\frac{\partial^2 \log|\boldsymbol{G}_j|}{\partial\psi_m\partial\psi_n} &= \frac{\partial}{\partial\psi_n}\left(\mathrm{tr}\left(\boldsymbol{G}_j^{-1}\frac{\partial \boldsymbol{G}_j}{\partial\psi_m}\right)\right)\\ &= \mathrm{tr}\left(\boldsymbol{G}_j^{-1}\frac{\partial^2 \boldsymbol{G}_j}{\partial\psi_m\partial\psi_n} - \boldsymbol{G}_j^{-1}\frac{\partial \boldsymbol{G}_j}{\partial\psi_n}\boldsymbol{G}_j^{-1}\frac{\partial \boldsymbol{G}_j}{\partial\psi_m}\right)\end{aligned}$$

The first and second derivatives of $\boldsymbol{G}_j$ are given in Equations (7.9)–(7.10).

Expectation

Since neither of Terms 1 and 3 contain random variables, it only remains to find the expectation of Term 2. If $\boldsymbol{x}$ is a vector of random variables with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{S}$, we have

$$\begin{aligned}\mathrm{E}(\boldsymbol{x}'\boldsymbol{A}) &= \boldsymbol{m}'\boldsymbol{A} \quad \text{and}\\ \mathrm{E}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) &= \mathrm{tr}(\boldsymbol{A}\boldsymbol{S}) + \boldsymbol{m}'\boldsymbol{A}\boldsymbol{m}\end{aligned}$$

$$\begin{aligned}\mathrm{E}\left(\frac{\partial^2 \boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j}{\partial\beta_m\partial\beta_n}\right) &= 2\boldsymbol{m}_j'\boldsymbol{A}_{1jmn} + \mathrm{tr}(\boldsymbol{A}_{2jmn}\boldsymbol{S}_j) + \boldsymbol{m}_j'\boldsymbol{A}_{2jmn}\boldsymbol{m}_j\\ \mathrm{E}\left(\frac{\partial^2 \boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j}{\partial\lambda_m\partial\lambda_n}\right) &= \mathrm{tr}(\boldsymbol{A}_{3jmn}\boldsymbol{S}_j) + \boldsymbol{m}_j'\boldsymbol{A}_{3jmn}\boldsymbol{m}_j\\ \mathrm{E}\left(\frac{\partial^2 \boldsymbol{e}_j'\boldsymbol{H}_j\boldsymbol{e}_j}{\partial\beta_m\partial\lambda_n}\right) &= -2\boldsymbol{m}_j'\boldsymbol{A}_{4jmn} + \mathrm{tr}(\boldsymbol{A}_{5jmn}\boldsymbol{S}_j) + \boldsymbol{m}_j'\boldsymbol{A}_{5jmn}\boldsymbol{m}_j\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{m}_j &= \mathrm{E}(\boldsymbol{e}_j)\\ \boldsymbol{S}_j &= \mathrm{Cov}(\boldsymbol{e}_j)\\ \boldsymbol{A}_{1jmn} &= \boldsymbol{H}_j\boldsymbol{T}_{2jmn} - \frac{\partial \boldsymbol{H}_j}{\partial\beta_n}\boldsymbol{X}_{j(m)} + \boldsymbol{T}_{1jm}'\boldsymbol{H}_j'\boldsymbol{X}_{j(n)} +\\ &\quad + \boldsymbol{T}_{1jn}'\boldsymbol{H}_j\boldsymbol{X}_{j(m)} - \frac{\partial \boldsymbol{H}_j}{\partial\beta_m}\boldsymbol{X}_{j(n)}\\ \boldsymbol{A}_{2jmn} &= 2\boldsymbol{H}_j\boldsymbol{T}_{3jmn} - 2\frac{\partial \boldsymbol{H}_j}{\partial\beta_n}\boldsymbol{T}_{1jm} + 2\boldsymbol{T}_{1jn}'\boldsymbol{H}_j\boldsymbol{T}_{1jm} -\\ &\quad - 2\frac{\partial \boldsymbol{H}_j}{\partial\beta_m}\boldsymbol{T}_{1jn} + \frac{\partial^2 \boldsymbol{H}_j}{\partial\beta_m\partial\beta_n}\\ \boldsymbol{A}_{3jmn} &= \frac{\partial^2 \boldsymbol{H}_j}{\partial\lambda_m\partial\lambda_n}\\ \boldsymbol{A}_{4jmn} &= \frac{\partial \boldsymbol{H}_j}{\partial\lambda_n}\boldsymbol{X}_{j(m)}\\ \boldsymbol{A}_{5jmn} &= \frac{\partial^2 \boldsymbol{H}_j}{\partial\beta_m\partial\lambda_n} - 2\frac{\partial \boldsymbol{H}_j}{\partial\lambda_n}\boldsymbol{T}_{1jm}\end{aligned}$$

Now

$$\begin{aligned} \mathrm{E}(\boldsymbol{e}_j) &= \mathrm{E}[\,\boldsymbol{W}_j^{-1}(\boldsymbol{y}_j - \boldsymbol{\pi}_j)] = \boldsymbol{W}_j^{-1}(\mathrm{E}(\boldsymbol{y}_j) - \boldsymbol{\pi}_j) \\ \mathrm{Cov}(\boldsymbol{e}_j) &= \mathrm{Cov}[\,\boldsymbol{W}_j^{-1}(\boldsymbol{y}_j - \boldsymbol{\pi}_j)] = \boldsymbol{W}_j^{-1}\mathrm{Cov}(\boldsymbol{y}_j)\,\boldsymbol{W}_j^{-1} \end{aligned}$$

where $\boldsymbol{\pi}_j = (\pi_{1j}, \ldots, \pi_{n_j j})'$.

$$\begin{aligned} \mathrm{E}(y_{ij}) &= \mathrm{E}[\mathrm{E}(y_{ij}|\boldsymbol{\delta}_j)] = \mathrm{E}[\pi_{ij}(\boldsymbol{\delta})] \\ &= \int_{\mathbb{R}^q} \cdots \int \pi_{ij}(\boldsymbol{\delta}_j) \frac{1}{(2\pi)^{q/2}} \exp\{-\frac{1}{2}\boldsymbol{\delta}_j'\boldsymbol{\delta}_j\}\mathrm{d}\boldsymbol{\delta}_j \end{aligned} \tag{7.11}$$

The (i,k)th element of $\mathrm{Cov}(\boldsymbol{y}_j)$ is given by $\mathrm{E}(y_{ij}y_{kj}) - \mathrm{E}(y_{ij})\mathrm{E}(y_{kj})$. If $i = k$,

$$\mathrm{E}(y_{ij}y_{kj}) = \mathrm{E}(y_{ij}^2) = \mathrm{E}(y_{ij}),$$

otherwise

$$\begin{aligned} \mathrm{E}(y_{ij}y_{kj}) &= \mathrm{E}[\mathrm{E}(y_{ij}y_{kj}|\boldsymbol{\delta}_j)] = \mathrm{E}[\pi_{ij}(\boldsymbol{\delta})\pi_{kj}(\boldsymbol{\delta})] \\ &= \int_{\mathbb{R}^q} \cdots \int \pi_{ij}(\boldsymbol{\delta}_j)\pi_{kj}(\boldsymbol{\delta}) \frac{1}{(2\pi)^{q/2}} \exp\{-\frac{1}{2}\boldsymbol{\delta}_j'\boldsymbol{\delta}_j\}\mathrm{d}\boldsymbol{\delta}_j. \end{aligned} \tag{7.12}$$

The integrals in Equations (7.11) and (7.12) may be calculated by quadrature (again, using the `quadl` function in MATLAB when $q = 1$).

7.2 Example: Dose-ranging trial

In this section, the problem of optimal design for a dose-ranging trial as described in Chapter 6 is considered. In this case, however, some of the parameters are allowed to vary randomly between subjects. The GLMM and its corresponding optimal designs are described below.

7.2.1 Model

Suppose that N individuals are given 2 doses (in sequence) of a drug which elicits a binary response (eg. success = 1, failure = 0). It is assumed that the effect of each dose carries over into the next period. The model is simplified slightly from the previous chapter, and the carryover effect of a dose is no longer assumed to be proportional to its direct effect.

Conditional on the j^{th} individual's parameters, the probability of success of its i^{th} dose can be modelled by the logistic regression model with $p = 3$ parameters,

$$\begin{aligned} \text{logit}\{P(y_{ij} = 1)|\boldsymbol{\theta}_j\} &= \theta_{j0} + \theta_{j1} d_{ij} + \theta_{j2} d_{(i-1)j} \\ &= \boldsymbol{x}_{ij}\boldsymbol{\theta}_j, \\ &\qquad i = 1, 2, \quad j = 1, \ldots, N, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{x}_{ij} &= (1,\, d_{ij},\, d_{(i-1)j}), \\ \boldsymbol{\theta}_j &= (\theta_{j0},\, \theta_{j1},\, \theta_{j2})', \end{aligned}$$

and where for $i \geq 1$, d_{ij} is the amount of the i^{th} dose given to the j^{th} patient, and $d_{0j} = 0$. θ_{j0} is referred to as the intercept parameter, θ_{j1} represents the direct effect of the dose, and θ_{j2} represents the effect of the dose carried over into the following period (the carryover effect). The vectors of parameters $\boldsymbol{\theta}_j$ are assumed to be independent random samples from a normal distribution,

$$\boldsymbol{\theta}_j \overset{iid}{\sim} \text{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

or alternatively $\boldsymbol{\theta}_j = \boldsymbol{\beta} + \boldsymbol{b}_j$ where

$$\boldsymbol{b}_j \overset{iid}{\sim} \text{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

This model assumes that every parameter varies randomly between subjects. A more general model may be formed by allowing only q of the p parameters to be random, and by treating the remaining parameters as fixed:

$$\begin{aligned} \text{logit}\{P(y_{ij} = 1)|\boldsymbol{\theta}_j\} &= \boldsymbol{x}_{ij}\boldsymbol{\theta}_j, \\ &= \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_j \\ &\qquad i = 1, 2, \quad j = 1, \ldots, N, \end{aligned}$$

where $\boldsymbol{b}_j \overset{iid}{\sim} \text{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is now a column vector of length $q \leq p$, and the $\boldsymbol{z}_{ij}$ (row vectors of length q) are the relevant subsets of the vectors $\boldsymbol{x}_{ij}$. This is equivalent to the model in Equation (7.1). Only the case with a single random parameter at a time is considered ($q = 1$)

in order to simplify calculation. A larger q will increase computation times considerably, due to the need for numerical evaluation of double or triple integrals. In practice, the intercept parameter is usually the only parameter assumed to vary between subjects in a pharmacodynamic model such as this (Yano *et al.*, 2001). Less frequently, the direct effect parameter is assumed to be random. Due to the single random parameter ($q = 1$) restriction imposed by computational constraints, only θ_{j0} and θ_{j1} will be considered random, as a random carryover effect would be highly unlikely without a random direct effect also in the model.

7.2.2 Optimal design

Define an elementary design for the j^{th} individual by the vector of doses $\boldsymbol{\xi}_j = (d_{1j}, d_{2j})'$. Both exact and approximate designs are considered, and are evaluated by using the D-optimality criterion, based on (a) Longford's approximation, $\boldsymbol{M}_L(\boldsymbol{\Psi}, \xi)$; (b) the more accurate approximation acknowledging the dependence of w_{ij} on $\boldsymbol{\beta}$, $\boldsymbol{M}_W(\boldsymbol{\Psi}, \xi)$; (c) the numerical method $\boldsymbol{M}_{Na}(\boldsymbol{\Psi}, \xi)$ based on the approximate log-likelihood; and (d) the numerical method involving numerical differentiation and integration, but no algebraic approximation, $\boldsymbol{M}_N(\boldsymbol{\Psi}, \xi)$. It is expected that methods (b) and (c) will produce very similar D-optimality criteria. In fact, (c) will only be used as to test that the complex criterion (b) is coded correctly.

As the numerical method $\boldsymbol{M}_N(\boldsymbol{\Psi}, \xi)$ is extremely computationally intensive (a single calculation of the information matrix for a very small design takes several minutes on a 2.4 GHz Pentium 4 PC), it will not be used for optimisation. Instead, it will only be used to compare efficiencies of designs generated by the first two methods, as it is assumed to be the most accurate representation of the information matrix available to us. As described in Section 7.1.1, this numerical method is only defined for exact designs, so even though both exact and approximate designs will be found using the two approximations, only the efficiencies of the exact designs will be compared.

All designs are locally optimum, with parameter means based on those used in the previous chapter:

$$\boldsymbol{\beta} = \mathrm{E}(\boldsymbol{\theta}_j) = (-1, 0.3, 0.25 \times 0.3)' = (-1, 0.3, 0.075)'.$$

7.2.3 Results

Before any optimal designs were found, a quick comparison of the four methods was performed. A 6-point exact design was chosen at random, and the D-optimality criterion was calculated using the four different information matrices. The intercept and direct effect parameters were assumed to be random, one at a time, and a range of variances of the random effects was used. Figure 7.1 shows the results. It can be seen that the 'less approximate' approximation to the information matrix ($\boldsymbol{M}_W$) gives almost exactly the same results as the more time-consuming numerical method using the approximate log-likelihood ($\boldsymbol{M}_{Na}$) as expected. Both of these are significantly closer to the 'gold standard' numerical method ($\boldsymbol{M}_N$) than Longford's approximation ($\boldsymbol{M}_L$), especially for larger variances. However, all three methods involving the approximation to the log-likelihood still seem to be fairly inaccurate when compared to $\boldsymbol{M}_N$, all over-estimating the criterion, and hence under-estimating the standard errors of parameter estimates on average. In light of this, the efficiencies of the exact optimal designs found with the computationally feasible algebraic methods ($\boldsymbol{M}_L$ and $\boldsymbol{M}_W$) will be compared in terms of the most accurate values of the criterion, calculated with $\boldsymbol{M}_N$.

Exact designs

Optimal exact designs are given for the two methods $\boldsymbol{M}_L$ and $\boldsymbol{M}_W$ in Tables 7.1 and 7.2, along with their criteria calculated by three different methods: $\boldsymbol{M}_L$, $\boldsymbol{M}_W$, and $\boldsymbol{M}_N$. When comparing two designs, the values of $|\boldsymbol{M}_N(\boldsymbol{\Psi}, \xi)|^{1/(p+q)}$ are compared since it is considered the closest to the true value of the D-optimality criterion. The ratio of these criteria give the efficiency of one design compared to another. Only 6-point designs are considered here due to the long computation times for some methods.

These optimal designs are compared to the optimal 6-point exact design for the fixed effects model, which has 4 points with $d_1 = 0$, $d_2 = 9.11$, and 2 points with $d_1 = 20$, $d_2 = 0$.

Table 7.1 shows D-optimal 6-point exact designs where the intercept parameter (θ_{j0}) varies randomly between subjects, and the other two parameters are fixed. For a 10% coefficient of variation (CV), the more accurate method ($\boldsymbol{M}_W$) actually produces a design

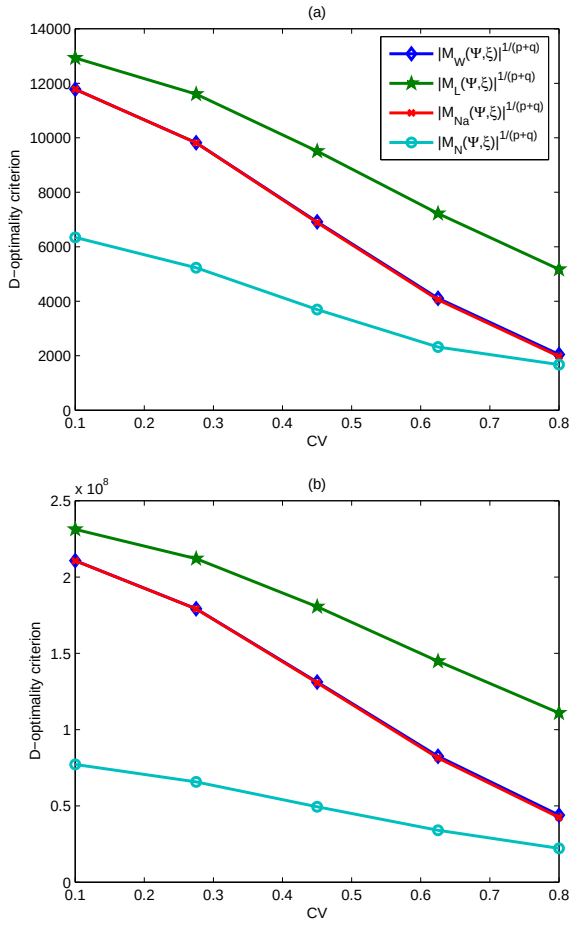

FIGURE 7.1: Comparison of different methods of calculation of the information matrix, for (a) θ_{j0} random, (b) θ_{j1} random.

which is 98.4% as efficient as Longford's method. The optimal exact design on 6 points for the fixed effects model is 93.2% as efficient as Longford's design. Similar patterns are seen as the CV is increased to 30% and 50%, with Longford's method producing the most efficient designs, followed by the more precise method, and with the fixed effects design being the least efficient. It is interesting to note that in this case the optimal design for the fixed effects model is always more than 90% as efficient as the most efficient method which incorporates the random effects, regardless of the size of the CV.

Table 7.2 shows the same information as in the previous table for the case where the direct effect parameter (θ_{j1}) varies randomly between subjects, and the other two parameters are

fixed. When the CV is 10%, the fixed effects design is actually about 4% *more* efficient than the most efficient design incorporating the random coefficient. These results are rather unexpected. The inclusion of the random coefficient in the design process is a substantially complex problem, but the exceptionally simpler information matrix for the fixed effects model produces a more efficient design in some cases. The large amount of error due to the Taylor expansion of the log-likelihood, as shown in Figure 7.1, may have a detrimental effect on the quality of designs found using this approximation. The fixed effects design does not perform as well for the larger values of CV, but is always more efficient than the designs found using the more precise approximation to the information matrix.

To check that these results are not specific to these parameter values, the designs are regenerated for a new set of parameter values, with larger direct and carryover effects:

$$\boldsymbol{\beta} = \mathrm{E}(\boldsymbol{\theta}_j) = (-1, 0.5, 0.25 \times 0.5)' = (-1, 0.5, 0.125)'.$$

The coefficient of variation is fixed at 30%. The results are shown in Table 7.3. The optimal 6-point exact fixed design for these parameters has 4 points at $d_1 = 0$, $d_2 = 5.58$, and 2 points at $d_1 = 20$, $d_2 = 0$.

When comparing the designs found using the two methods incorporating the random coefficients, it can be seen that the designs found using the more precise approximation again produces slightly less efficient designs. The fixed effects design also produces the most efficient design in one of the two cases.

Approximate designs

Optimal approximate designs are given in Tables 7.4 and 7.5 and in Figures 7.2 and 7.3. The approximate fixed effects design from the previous chapter is also given as a plot in Figure 7.4. As for the plots of the approximate designs in Chapter 5, each support point is represented by its position in the xy-plane, and its corresponding weight is proportional to the size (area) of the plotted point.

From Figure 7.2 it can be seen that the size of the CV makes very little difference to the positions and weights of the support points when the intercept parameter is random. Although for the $\boldsymbol{M}_L$ method, the smaller design point actually disappears altogether when

Table 7.1: Exact optimal designs when the intercept parameter (θ_{j0}) is random.

	Optimisation			Criterion		
CV	Method	d_1	d_2	$\lvert\boldsymbol{M}_L\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_W\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_N\rvert^{1/(p+q)}$
10%	$\boldsymbol{M}_L$	0.00	6.67	6.9148	–	5.6228
		0.00	6.67			
		0.00	6.67			
		0.00	6.67			
		20.00	0.00			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	5.08	–	6.5103	5.5333
		0.00	5.08			
		0.00	5.08			
		8.45	0.00			
		20.00	0.00			
		20.00	0.00			
	Fixed effects design			–	–	5.2431
30%	$\boldsymbol{M}_L$	0.00	6.76	6.7324	–	5.3599
		0.00	6.76			
		0.00	6.76			
		0.00	6.76			
		20.00	0.00			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	5.06	–	6.2209	5.2577
		0.00	5.06			
		0.00	5.06			
		9.06	0.00			
		20.00	0.00			
		20.00	0.00			
	Fixed effects design			–	–	5.0483
50%	$\boldsymbol{M}_L$	0.00	6.91	6.3981	–	4.9049
		0.00	6.91			
		0.00	6.91			
		0.00	6.91			
		20.00	0.00			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	5.08	–	5.6863	4.7865
		0.00	5.08			
		0.00	5.08			
		10.20	0.00			
		20.00	0.00			
		20.00	0.00			
	Fixed effects design			–	–	4.6927

Table 7.2: Exact optimal designs when the direct effect parameter (θ_{j1}) is random.

	Optimisation			Criterion		
CV	Method	d_1	d_2	$\lvert\boldsymbol{M}_L\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_W\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_N\rvert^{1/(p+q)}$
10%	$\boldsymbol{M}_L$	0.00	0.00	37.8542	–	34.9416
		0.00	9.08			
		8.82	7.89			
		8.82	7.89			
		8.82	7.89			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	0.00	–	27.0181	32.7066
		6.37	6.04			
		6.37	6.04			
		6.37	6.04			
		6.37	6.04			
		20.00	0.00			
	Fixed effects design			–	–	36.3835
30%	$\boldsymbol{M}_L$	0.00	0.00	35.1658	–	42.8165
		0.00	9.10			
		8.51	8.33			
		8.51	8.33			
		20.00	0.00			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	0.00	–	23.1245	31.2136
		5.66	5.39			
		5.66	5.39			
		5.66	5.39			
		5.66	5.39			
		20.00	0.00			
	Fixed effects design			–	–	38.5235
50%	$\boldsymbol{M}_L$	0.00	9.17	31.4545	–	28.5358
		0.00	9.17			
		0.00	9.17			
		8.60	8.27			
		20.00	0.00			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	4.09	–	17.5568	16.2869
		5.28	20.00			
		5.28	20.00			
		5.70	0.00			
		5.70	0.00			
		5.72	3.59			
	Fixed effects design			–	–	26.8325

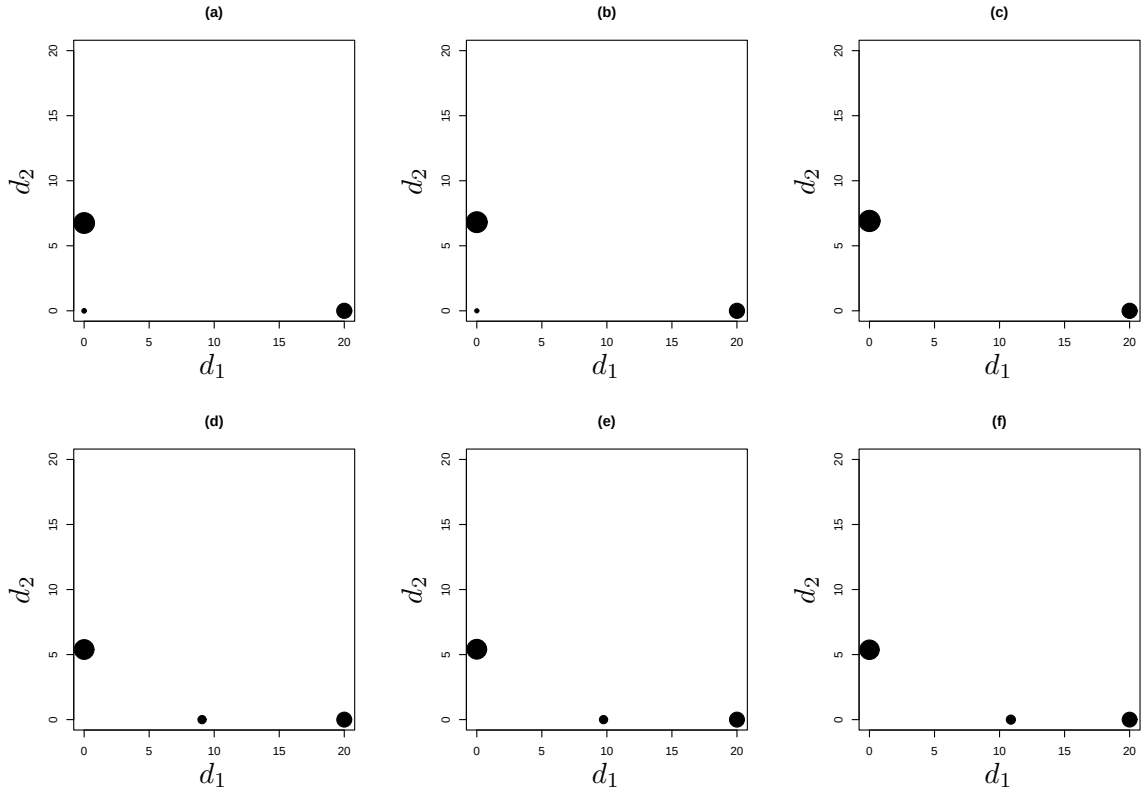

FIGURE 7.2: Approximate optimal designs when the intercept parameter (θ_{j0}) is random, using method $\boldsymbol{M}_L$ with (a) CV = 10%, (b) CV = 30%, (c) CV = 50%; and using method $\boldsymbol{M}_W$ with (d) CV = 10%, (e) CV = 30%, (f) CV = 50%.

the CV is increased to 50%. The difference between the designs generated by the $\boldsymbol{M}_L$ and $\boldsymbol{M}_W$ methods is minor, and all are fairly similar to the fixed effects design, with the major exception of the (0,0) point being moved to a higher first dose for the $\boldsymbol{M}_W$ method.

The designs given in Figure 7.3 for a random direct effect parameter show a marked difference in the two methods of design construction, as well as a more significant effect of the size of CV than in the case of a random intercept. All of these designs are also significantly different from the fixed effects design. The only support point in common between all 7 designs in Figures 7.3 and 7.4 is (20,0). The variability of designs appear to reflect the high variability in criterion values of the exact designs for this random parameter given in Table Table 7.2.

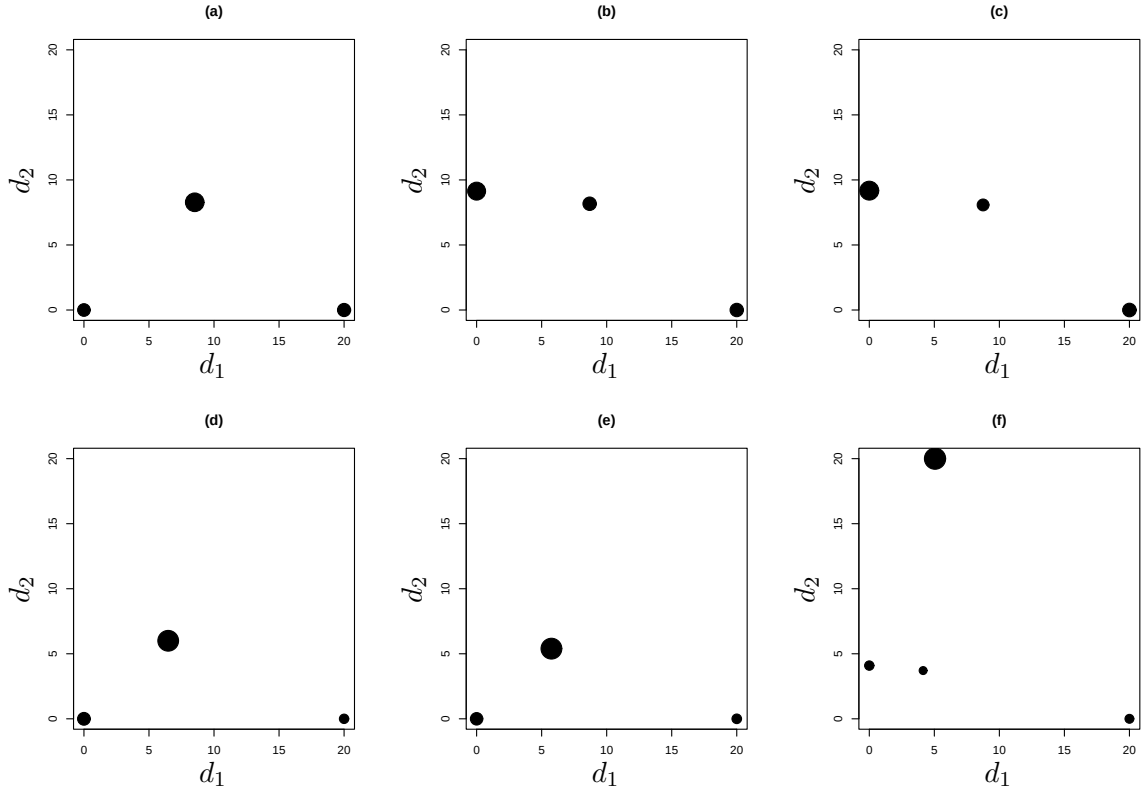

Figure 7.3: Approximate optimal designs when the direct effect parameter (θ_{j1}) is random, using method $\boldsymbol{M}_L$ with (a) CV = 10%, (b) CV = 30%, (c) CV = 50%; and using method $\boldsymbol{M}_W$ with (d) CV = 10%, (e) CV = 30%, (f) CV = 50%.

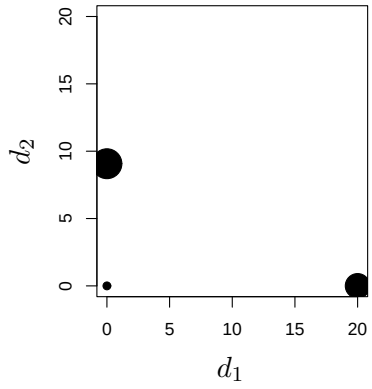

Figure 7.4: Approximate optimal design for the fixed effects model.

7.3 Discussion

In this chapter, two methods of computing optimal designs for logistic regression models with random coefficients were considered. Both involved the construction of the information matrix by using a Taylor expansion of the log-likelihood. One of the methods involved further simplifications and assumptions, the other method did not.

The information matrices calculated by these two methods were compared to a more computationally intensive numerical method, assumed to be an accurate representation of the true information matrix. It was shown that for these models, the method with no further simplifications and assumptions produced more accurate information matrices than the more approximate method, as expected.

Even though the much more complex method, where the only simplification was the use of the Taylor expansion, was shown to give information matrices significantly closer to the true values, the optimal designs based on these methods were found in all cases to be slightly less efficient than designs found using the much simpler, quicker method.

In fact, the use of the simpler fixed effects model produced designs which were in most cases almost as efficient as the designs incorporating the random effects. When the direct effect parameter in the model is allowed to vary between subjects, the fixed effects design is actually sometimes slightly more efficient than the designs found using the more sophisticated methods. In light of these results, it may not be worth incorporating random coefficients into the design process at all until more accurate methods of constructing the information matrix are developed.

The reason for the unexpected low efficiencies for the design optimisation methods incorporating random effects is not immediately apparent, and may be a topic of future research in this area.

However, rather than looking for a more accurate information matrix for these models, it may be more beneficial to look to alternative models for this data which simplifies the algebra. Two such alternatives currently under consideration are: the probit link function for binary data, which produces a very similar response curve to the logit link function used

here, but which appears to be somewhat simpler to work with algebraically; and the beta-binomial model which, after some preliminary investigation, may have the same advantage.

Table 7.3: Exact optimal designs for a different set of parameter values, with CV = 30%.

Random	Optimisation			Criterion		
Coefficient	Method	d_1	d_2	$\lvert\boldsymbol{M}_L\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_W\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_N\rvert^{1/(p+q)}$
θ_{j0}	$\boldsymbol{M}_L$	0.00	4.32	4.3425	–	3.5382
(intercept)		0.00	4.32			
		0.00	0.77			
		5.16	0.00			
		5.16	0.00			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	2.37	–	4.0001	3.4771
		0.00	2.37			
		3.78	0.00			
		5.64	0.00			
		5.64	0.00			
		20.00	0.00			
	Fixed effects design			–	–	3.3159
θ_{j1}	$\boldsymbol{M}_L$	0.00	5.58	14.0018	–	14.7351
(direct effect)		0.00	5.58			
		0.00	5.58			
		5.24	4.83			
		5.24	4.83			
		20.00	0.00			
	$\boldsymbol{M}_W$	0.00	0.00	–	10.0803	14.1752
		0.00	3.30			
		3.59	3.49			
		3.59	3.49			
		3.59	3.49			
		20.00	0.00			
	Fixed effects design			–	–	18.1463

Table 7.4: Approximate optimal designs when the intercept parameter (θ_{j0}) is random.

	Optimisation				Criterion	
CV	Method	d_1	d_2	w	$\lvert\boldsymbol{M}_L\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_W\rvert^{1/(p+q)}$
10%	$\boldsymbol{M}_L$	0.00	0.00	0.0312	1.1529	–
		0.00	6.75	0.6327		
		20.00	0.00	0.3361		
	$\boldsymbol{M}_W$	0.00	5.39	0.5723	–	1.0862
		9.05	0.00	0.0995		
		20.00	0.00	0.3282		
30%	$\boldsymbol{M}_L$	0.00	0.0	0.0250	1.1224	–
		0.00	6.82	0.6382		
		20.00	0.00	0.3368		
	$\boldsymbol{M}_W$	0.00	5.41	0.5719	–	1.0380
		9.74	0.00	0.0995		
		20.00	0.00	0.3286		
50%	$\boldsymbol{M}_L$	0.00	6.91	0.6624	1.0664	–
		20.00	0.00	0.3376		
	$\boldsymbol{M}_W$	0.00	5.37	0.5548	–	0.9487
		10.87	0.00	0.1227		
		20.00	0.00	0.3226		

Table 7.5: Approximate optimal designs when the direct effect parameter (θ_{j1}) is random.

Optimisation					Criterion	
CV	Method	d_1	d_2	w	$\lvert\boldsymbol{M}_L\rvert^{1/(p+q)}$	$\lvert\boldsymbol{M}_W\rvert^{1/(p+q)}$
10%	$\boldsymbol{M}_L$	0.00	0.00	0.2350	6.4381	–
		8.52	8.28	0.5098		
		20.00	0.00	0.2552		
	$\boldsymbol{M}_W$	0.00	0.00	0.2393	–	4.5958
		6.48	6.00	0.6305		
		20.00	0.00	0.1302		
30%	$\boldsymbol{M}_L$	0.00	9.14	0.4719	5.9436	–
		8.69	8.17	0.2655		
		20.00	0.00	0.2626		
	$\boldsymbol{M}_W$	0.00	0.00	0.2283	–	3.9108
		5.75	5.39	0.6361		
		20.00	0.00	0.1356		
50%	$\boldsymbol{M}_L$	0.00	9.18	0.5237	5.2801	–
		8.75	8.08	0.2056		
		20.00	0.00	0.2707		
	$\boldsymbol{M}_W$	0.00	4.09	0.1290	–	4.1079
		4.14	3.71	0.0928		
		5.05	20.00	0.6618		
		20.00	0.00	0.1164		

Chapter 8

Conclusion

This thesis has described a number of recent developments in optimal design of experiments for nonlinear and generalised linear models, with applications to real-life and hypothetical examples arising in pharmacology and other areas. While the application of these techniques has been novel in itself, this thesis has also presented new methods of designing experiments with multiple objectives (model discrimination and parameter estimation), as well as designs for generalised linear models with random effects.

Product optimal designs for two competing nonlinear models were shown to be efficient for estimation of parameters in both models, but they can be quite poor for discriminating between the two models according to the T-optimality criterion, particularly when the number of support points is small. The use of both conditional and hybrid designs allow product optimal designs to be augmented with additional support points from T-optimal designs (or points close to the T-optimal designs in the case of conditional designs), which can increase the T-efficiencies significantly without having a great impact on the D-efficiencies. The converse is seen when starting with a T-optimal design and adding support points from the product optimal design. The hybrid designs seem to be very similar to the conditional designs in terms of support points/weights and T- and D-efficiencies, so the use of these designs would be preferable to the conditional designs, as the construction of the hybrid designs requires no further optimisation once the T-optimal and product optimal designs have been found.

Although the T-optimality criterion provides a measure of dissimilarity of the two models at the design points, a better approach to the evaluation of the designs in terms of model discrimination may be to conduct power tests by simulation in a similar manner to the methods used for the calculation of power for generalised linear models given in this thesis. It would be interesting to examine the relationship between the T-optimality criterion and the power found by these simulations. This is an area for future investigation. Further topics of research could also include the evaluation of the use of conditional and hybrid designs for discrimination between more than two nonlinear models, with the additional goal of efficient parameter estimation under each model. Atkinson and Fedorov (1975b) have defined the T-optimality criterion for such a situation. The methods of construction of conditional and hybrid designs described in this thesis are general enough to apply to alternative criteria such as this.

The product criterion for parameter estimation in nonlinear models was applied to a real-life design optimisation problem in pharmacokinetics, in which the aim was to discriminate between two candidate models as well as to estimate the parameters efficiently under each model. The models involved were rather complex, involving between subject variability and multiple responses, and the solution to the differential equations describing the model for nonlinear elimination does not exist in closed form. Numerical solutions to these equations were able to be incorporated into the design process, as well as the random effects and multiple responses. The resulting product criterion was shown by simulations to produce a design which is efficient in terms of both model discrimination and parameter estimation. This is the third published case of the prospective use of optimal design for a pharmacokinetic experiment (the first involving nonlinear elimination), and illustrates the usefulness of optimal design as opposed to more traditional empirical methods of design in this area. Some ideas investigated in that work are now being adopted by a major international pharmaceutical company for the design of their pharmacokinetic studies.

The experiment itself was constrained by the experimenters in a number of aspects, but the design process was also under time constraints. In the limited time available, the methods were researched and applied to these models, and the computationally intensive optimisation was carried out. Given more time and resources, a more sophisticated method

of incorporating multiple responses for mixed effects models could be considered, although the simplifications and assumptions made here seem to be fairly inconsequential, as the standard errors predicted by the approximated information matrix are acceptably close to those estimated from simulated data. The use of numerical derivatives could also be replaced by a more elegant method such as the 'direct method' described in Atkinson and Bogacka (2002).

The method of constructing hybrid designs for nonlinear models was adapted to designs for generalised linear models, where again the objective of the experiment is to discriminate between two models (nested in this case), and to efficiently estimate the parameters under each model. Conditional designs were not used here as the differences between conditional and hybrid designs appeared to be minimal for nonlinear models, and the construction of conditional designs is much more computationally intensive. Similar trends were seen for GLMs as for nonlinear models when using hybrid designs, with a reasonable trade-off between estimation and discrimination. In addition to the comparison of T- and D-optimality criteria, these hybrid designs were evaluated by power tests involving large simulations. While the difference in power between the product optimal and T-optimal designs is often minimal, a distinct trend of increasing power is seen as the hybrid designs place more weight on the T-optimal design points, and vice versa. This is encouraging, as it shows that the T-optimality criterion does actually produce designs which are more effective than others in terms of model discrimination.

While it has been seen that the T-optimality criterion value of a design increases with its power to correctly discriminate between two models in these cases, the actual relationship between the two is unknown. Given the information matrix for a design, the standard errors of parameter estimates can be predicted by the diagonal elements of the inverse of the matrix. It would be beneficial to approximate the power of a design (very computationally expensive to compute using simulations) in a manner analogous to this, based on the design's T-optimality criterion. As for nonlinear models, the construction of hybrid designs for GLMs described in this thesis can be generalised to apply to three or more models, using the T-optimality criterion for these situations as described by Ponce de Leon and Atkinson (1992), and the product of D-optimality criteria for all models for the product design. These

techniques are also easily applied to models with more than one or two explanatory variables, not generally addressed in research in optimal design for GLMs. The T_E-optimality criterion outlined in this thesis has more attractive distributional qualities than the T-optimality criterion, although its optimisation is at present a computationally intractable problem. Further work on this or a related criterion may lead to methods of producing designs more effective for model discrimination.

To complement the real-life example of optimal design for a nonlinear pharmacokinetic model, the optimal design of a hypothetical pharmacodynamic model involving a logistic regression model was considered. Optimisation of crossover and parallel (single period) designs were shown to offer a significant improvement over more traditionally used balanced design with an increase in parsimony due to the significant decrease in the number of support points. Although the designs found are locally optimal, sensitivity analyses show that these designs are still quite efficient if the parameter values are misspecified to a reasonably large degree. Designs which are composites of optimal crossover and parallel designs afford experimenters greater flexibility in such trials without adversely affecting the efficiency of parameter estimates.

It would be interesting to extend these methods to account for crossover designs with more than two periods. Although a carryover effect of greater than one period is not generally considered in these models in most situations, second-order (or higher) carryover effects into the models may be incorporated into the design process. Optimal designs under these conditions may or may not have the same properties as those presented in this thesis.

Generalised linear mixed models have been neglected in the optimal design literature, and the work presented in this thesis highlights the enormous technical and computational difficulties associated with design optimisation for such models. Two approximations to the information matrix for a logistic regression model were given, with one, $\boldsymbol{M}_W$, involving less simplification than the other, $\boldsymbol{M}_L$, but both relying on a Taylor expansion of the log-likelihood function. Using a computationally intensive numerical method, $\boldsymbol{M}_N$, to compare designs found using the two approximate methods, the optimal designs based on $\boldsymbol{M}_W$, the matrix using less simplification (and much greater algebraic complexity), was found to be slightly less efficient than designs produced by $\boldsymbol{M}_L$. Designs optimised using the much

simpler information matrix which ignored the random effects altogether were actually found in some cases to be more efficient than the two methods which incorporated the random effects.

The reasons for the unexpected differences in efficiency between these methods remain to be understood, and may be investigated further. A more profitable approach, however, may be to focus instead on alternative models such as the beta-binomial model or the same model used here with the logit link function replaced by the probit link function, which has a similar response curve. Early exploration has shown that these models may be more algebraically manageable. Once issues with the information matrix have been adequately dealt with, the crossover designs could be expanded as described above to include more periods and higher-order carryover effects.

Another aim of the design of an experiment not mentioned in this thesis, which may be addressed by optimal design, is that of reducing the bias of parameter estimates in nonlinear and generalised linear models. This is a significant area of future research. Optimal design for mixed effects models in general also require further investigation. The fixed and random effects may be treated separately in a type of compound criterion, or in the case where random effects are treated as nuisance parameters, we may adjust for them through the use of D_s-optimal designs. Bayesian methods for experimental design, such as those outlined in Chaloner and Verdinelli (1995) and Clyde (2001), are also being considered for many problems discussed in this thesis.

Appendix A

Appendix: Source code

This appendix contains MATLAB code (M-files) used to find the optimal designs given in this thesis. The search algorithms described in Section 2.4, simulated annealing and the cross-entropy method, are given first, followed by some miscellaneous additional functions used by these algorithms.

A.1 anneal.m: Simulated annealing algorithm

```
function [X,FVAL,elapsedtime,EXITFLAG,OUTPUT,LAMBDA,GRAD,HESSIAN] = ...
    anneal(FUN,X,LB,UB,options,varargin)
%ANNEAL Finds the constrained maximum of a function of several variables
% by simulated annealing.
%
%   ANNEAL solves problems of the form:
%       max F(X)  subject to:  LB <= X <= UB
%
%   X=ANNEAL(FUN,X0,LB,UB) starts at X0 and finds a maximum X to the
%   function FUN, subject to the lower and upper bounds LB and UB. FUN
%   accepts input X and returns a scalar function value F evaluated at X.
%   X0 may be a scalar, vector, or matrix.
%
%   X=ANNEAL(FUN,X0,LB,UB,OPTIONS) maximises with the default
%   optimisation parameters replaced by values in the structure OPTIONS.
%   Use OPTIONS = [] as a place holder if no options are set.
%   The list of options available:
%
%     initprob        initial probability of accepting downhill steps
%                     (default = 0.95)
```

```
%     cool            cooling rate: temp = cool*temp
%                     (default = 0.9)
%     ncyct           number of cycles between changes in temperature
%                     (default = 20)
%     ncycs           number of cycles between changes in step size
%                     (default = 10)
%     ncyci           number of initial cycles (to calculate temperature)
%                     (default = 100)
%     vmin            minimum change in step size
%                     (default = 1e-3)
%     maxit           maximum number of iterations
%                     (default = 1e7)
%     dispint         number of temperature changes between displaying
%                     interim results (default = 5)
%     Vstep           initial step size
%                     (default = UB-LB)
%     weighted        weighted = 1: constrain last column of X to sum to 1
%                     (default = 1)
%
%
%   X=ANNEAL(FUN,X0,LB,UB,OPTIONS,P1,P2,...) passes the
%   problem-dependent parameters P1,P2,... directly FUN:
%    feval(FUN,X,P1,P2,...).
%   Pass empty matrices for OPTIONS, LB, and UB to use the
%   default values.
%
%   [X,FVAL]=ANNEAL(FUN,X0,...) returns the value of the objective
%   function FUN at the solution X.
%
%   Examples
%     FUN can be specified using @:
%        X = anneal(@humps,...)
%     In this case, F = humps(X) returns the scalar function value F of
%     the HUMPS function evaluated at X.
%
%     FUN can also be an inline object:
%        X = anneal(inline('3*sin(x(1))+exp(x(2))'),[1;1],[0;0],[1;1])
%     returns X = [0;0].
%
%     Calling with options:
%        options = struct('vmin',1e-5,'ncyct',20,'ncycs',60,...
%                         'weighted',1,'dispint',1);
%        X = anneal(@fun,x0,lb,ub,options)
%
```

```
%   Copyright 2005 Tim Waterhouse, all rights reserved

if nargin < 4, error('ANNEAL requires at least four input arguments'); end
if nargin < 5
    options = struct('initprob',0.95,'cool',0.9,'ncyct',10,'ncycs',20,...
 'ncyci',100,'vmin',1e-3,'maxit',1e7,'dispint',5,'Vstep',UB-LB,'weighted',1);
else
    % initial probability of accepting downhill steps
    if ~isfield(options,'initprob')
        options.initprob = 0.95;
    end
    % cooling rate: temp = cool*temp
    if ~isfield(options,'cool')
        options.cool = 0.9;
    end
    % number of cycles between changes in temperature
    if ~isfield(options,'ncyct')
        options.ncyct = 10;
    end
    % number of cycles between changes in step size
    if ~isfield(options,'ncycs')
        options.ncycs = 20;
    end
    % number of initial cycles (to calculate temperature)
    if ~isfield(options,'ncyci')
        options.ncyci = 100;
    end
    % minimum change in step size (stop when this value is reached)
    if ~isfield(options,'vmin')
        options.vmin = 1e-3;
    end
    % maximum number of iterations
    if ~isfield(options,'maxit')
        options.maxit = 1e7;
    end
    % number of temperature changes between displaying interim results
    if ~isfield(options,'dispint')
        options.dispint = 5;
    end
    % initial step size
    if ~isfield(options,'Vstep')
        options.Vstep = UB-LB;
    end
    % weighted = 1: constrain last column of X to sum to 1
```

```
    if ~isfield(options,'weighted')
        options.weighted = 1;
    end
end

[m,n]=size(X);

Vstep = options.Vstep;
b = -inf;
start = clock;
cr=0;               % counter for number of iterations
crt=0;              % counter for number of tempe changes
RJCT=zeros(m,n);    % vector of number of rejects per ncycs iterations

if options.weighted
    X(:,n) = X(:,n)./sum(X(:,n));
end
bestX = X;
d = feval(FUN,X,varargin{:});
opt = d;
best = opt;
cnt = 0;

umax = 0; omin = inf; omax = best;
for i = 1:options.ncyci
    cnt = cnt+1;
    for j=1:m
        for k = 1:n
            tmpX = -ones(m,n)*Inf;
            tmpX = X;
            tmpX(j,k) = min(max(tmpX(j,k) + (rand*2-1)*Vstep(j,k), ...
                LB(j,k)),UB(j,k));
            if options.weighted
                if k == n
                    tmpX(:,n) = tmpX(:,n)./sum(tmpX(:,n));
                end
            end
            d1 = feval(FUN,tmpX,varargin{:});
            if d1 == -Inf
                d1 = d;
            end
            if abs(d - d1) > umax
                umax = abs(d - d1);
            end
```

```
            if d1 > omax
                omax = d1;
                bestX = tmpX;
            end
            if d1 < omin
                omin = d1;
            end
        end
    end
    if rem(cnt,options.dispint)==0
        disp([cnt omax])
    end
end

[umax, omax, omin, omax-omin]
best = omax;
opt = omax;
if umax == 0 | umax == Inf
    temp = 1e3
else
    temp = -umax/log(options.initprob)
end
cnt = 0;
bestX

save interim.mat

while max(max(Vstep./(UB-LB))) > options.vmin & cr < options.maxit
    % restart with current best
    X = bestX;
    for cyct = 1:options.ncyct
        for cycs = 1:options.ncycs
            for j=1:m
                for k = 1:n
                    cr = cr + 1;
                    tmpX = X;
                    if size(X) ~= [m,n]
                        X
                    end
                    tmpX(j,k) = X(j,k) + (rand*2-1)*Vstep(j,k);

                    if tmpX(j,k) < LB(j,k) | tmpX(j,k) > UB(j,k)
                        d = -Inf;
                    else
```

```
                    if options.weighted
                        if k == n
                            tmpX(:,n) = tmpX(:,n)./sum(tmpX(:,n));
                        end
                    end
                    d = feval(FUN,tmpX,varargin{:});
                end
                if d > opt
                    X = tmpX;
                    opt = d;
                else
                    if exp((d-opt)/temp)>rand
                        X = tmpX;
                        opt = d;
                    else
                        RJCT(j,k) = RJCT(j,k) + 1;
                    end
                end
                if opt > best
                    bestX = X;
                    best = opt;
                end
            end
        end
    end
    FRJCT = max(RJCT/options.ncycs,0.01);
    c_sa = min(temp,0.01);
    Vstep=Vstep./(FRJCT/.5)+c_sa;
    RJCT = zeros(m,n);
end

if options.weighted
    [I,J]=ind2sub(size(bestX(:,n)),...
        find(bestX(:,n)<options.vmin & Vstep(:,n)<options.vmin));
    if length(I) > 0
        tmpX = bestX;
        tmpX(I,:)=[];
        tmpX(:,n) = tmpX(:,n)./sum(tmpX(:,n));
        d = feval(FUN,tmpX,varargin{:});
        if d >= best
            bestX = tmpX;
            best = d;
            Vstep(I,:)=[];
            RJCT(I,:)=[];
```

```
                LB(I,:)=[];
                UB(I,:)=[];
                m = m-length(I);
            end
        end
    end

    temp = temp*options.cool;
    cnt = cnt+1;
    if rem(cnt,options.dispint)==0
        disp([temp mean(mean(FRJCT)) max(max(Vstep./(UB-LB)))])
        disp(bestX)
        best
    end
    save interim.mat
    if options.weighted & m==1
        break
    end
end

if cr >= options.maxit
    disp('Maximum number of iterations reached')
else
    disp('Cooling finished.')
end
disp('Elapsed time:')
elapsedtime = etime(clock,start)
disp('Number of temperature changes:')
disp(cnt)

if options.weighted
    X = sortrows(collapse(bestX,options.vmin))
else
    X = sortrows(bestX)
end
FVAL = best

save interim.mat
```

A.2 crossentropy.m: The cross-entropy method

```
function [bestx,bestcrit,elapsedtime] = ...
        crossentropy(FUN,LB,UB,options,varargin)
%CROSSENTROPY Finds the constrained maximum of a function of several
```

```
% variables by the cross-entropy method.
%
%   CROSSENTROPY solves problems of the form:
%       max F(X)  subject to:  LB <= X <= UB
%
% FUN = F(X), function containing the criterion to be maximised
% LB = Lower bound on the design matrix
% UB = Upper bound on the design matrix
%
% Copyright 2005 Tim Waterhouse, all rights reserved

start = clock;

if nargin < 3, error('CE requires at least three input arguments'); end

if nargin == 3
    options = struct('tol',1e-2,'rho',0.1,...
        'alpha1',0.9,'alpha2',0.3,'N',1000,'weighted',1,'dispint',10,...
        'num_inject',3,'h_inject',2);
else
    if ~isfield(options,'tol')
        options.tol = 1e-2;
    end
    if ~isfield(options,'rho')
        options.rho = 0.1;
    end
    if ~isfield(options,'alpha1')
        options.alpha1 = 0.9;
    end
    if ~isfield(options,'alpha2')
        options.alpha2 = 0.3;
    end
    if ~isfield(options,'N')
        options.N = 1000;
    end
    if ~isfield(options,'weighted')
        options.weighted = 1;
    end
    if ~isfield(options,'dispint')
        options.dispint = 10;
    end
    if ~isfield(options,'num_inject')
        options.num_inject = 3;
    end
```

```
    if ~isfield(options,'h_inject')
        options.h_inject = 2;
    end
end

[r,c] = size(LB);
LB = reshape(LB,r*c,1);
UB = reshape(UB,r*c,1);

% initial estimates of mu and sigma
mu = LB + (UB-LB)/2;
sigma = (UB-LB)/4;

bestx = mu;
bestcrit = feval(FUN,reshape(mu,r,c),varargin{:});
currbest = bestcrit;
inject = 0;
t=0;
currbestx=[];

while inject <= options.num_inject
    t=t+1;
    x=[];
    critvec=[];
    for s=1:options.N
        tmp = normt_rnd(mu,sigma.^2,LB,UB);
        if options.weighted
            k=sum(tmp(r*(c-1)+1:r*c));
            tmp(r*(c-1)+1:r*c)=tmp(r*(c-1)+1:r*c)/k;
        end
        critvec = [critvec, feval(FUN,reshape(tmp,r,c),varargin{:})];
        x = [x, tmp];
    end
    [critsort,ind]=sort(critvec);
    ind = ind(ceil(options.N*(1-options.rho)):options.N);
    currbestx=[];
    for i = ind
        currbestx = [currbestx, x(:,i)];
    end
    mu = options.alpha1*mean(currbestx,2) + (1-options.alpha1)*mu;
    sigma = options.alpha2*std(currbestx,1,2) + (1-options.alpha2)*sigma;
    q = max(sigma);
    prevbest = currbest;
    currbest = max(critvec);
```

```
    if q <= options.tol
        inject = inject + 1;
        sigma = sigma + abs(currbest-prevbest)*options.h_inject;
        q = max(sigma);
    end
    if currbest > bestcrit
        bestx = x(:,ind(options.rho*options.N));
        bestcrit = currbest;
    end
    if rem(t,options.dispint)==0
        reshape(bestx,r,c)
        [bestcrit,q]
    end
    bestx = reshape(bestx,r,c);
    mu = reshape(bestx,r,c);
    sigma = reshape(sigma,r,c);
    LB = reshape(LB,r,c);
    UB = reshape(UB,r,c);
    if options.weighted
        [I,J]=ind2sub(size(bestx(:,c)),...
            find(bestx(:,c)<options.tol & sigma(:,c)<options.tol));
        if length(I) > 0
            tmpx = bestx;
            tmpx(I,:)=[];
            tmpx(:,c) = tmpx(:,c)./sum(tmpx(:,c));
            d = feval(FUN,tmpx,varargin{:});
            if d >= bestcrit
                bestx = tmpx;
                bestcrit = d;
                mu(I,:)=[];
                sigma(I,:)=[];
                LB(I,:)=[];
                UB(I,:)=[];
                r = r-length(I);
            end
        end
    end
    bestx = reshape(bestx,r*c,1);
    mu = reshape(bestx,r*c,1);
    sigma = reshape(sigma,r*c,1);
    LB = reshape(LB,r*c,1);
    UB = reshape(UB,r*c,1);

end
```

```
if options.weighted
    bestx=sortrows(collapse(reshape(bestx,r,c),options.tol));
else
    bestx=sortrows(reshape(bestx,r,c));
end
elapsedtime = etime(clock,start);
```

A.3 Miscellany

A.3.1 collapse.m

The following function 'collapses' an approximate design to its simplest form. That is, it combines equal (or close to equal) support points into a single support point with a weight equal to the sum of the weights of its constituent points. It also removes any points with zero weight (or close to zero weight).

```
function Xnew = collapse(X,tol)
% Collapses an approximate design to its simplest form

[n,m] = size(X);

% Combine identical (or nearly identical) support points
for guava = 1:(n-1)
    for cumquat = (guava+1):n
        if sum(abs((X(guava,1:(m-1))-X(cumquat,1:(m-1))))) < tol
            X(guava,m) = X(guava,m) + X(cumquat,m);
            X(cumquat,m) = 0;
        end
    end
end

% Remove support points with nearly zero weight
[I,J] = find(X(:,m)<tol);
X(I,:) = [];

Xnew = X;
```

A.3.2 normt_rnd.m

The `normt_rnd` function draws a random sample from a truncated normal distribution. This M-file is not original work, and was downloaded from the following location:

```
http://www.spatial-econometrics.com/distrib/normlt_rnd.m

function result = normt_rnd(mu,sigma2,left,right)
% PURPOSE: random draws from a normal truncated to (left,right) interval
% ---------------------------------------------------
% USAGE: y = normt_rnd(mu,sigma2,left,right)
% where:   mu = mean (nobs x 1)
%      sigma2 = variance (nobs x 1)
%        left = left truncation points (nobs x 1)
%       right = right truncation points (nobs x 1)
% ---------------------------------------------------
% RETURNS: y = (nobs x 1) vector
% ---------------------------------------------------
% NOTES: use y = normt_rnd(mu,sigma2,left,mu+5*sigma2)
%        to produce a left-truncated draw
%        use y = normt_rnd(mu,sigma2,mu-5*sigma2,right)
%        to produce a right-truncated draw
% ---------------------------------------------------
% SEE ALSO: normlt_rnd (left truncated draws), normrt_rnd (right truncated)
%

% adopted from Bayes Toolbox by
% James P. LeSage, Dept of Economics
% University of Toledo
% 2801 W. Bancroft St,
% Toledo, OH 43606
% jpl@jpl.econ.utoledo.edu

% For information on the Bayes Toolbox see:
% Ordinal Data Modeling by Valen Johnson and James Albert
% Springer-Verlag, New York, 1999.



if nargin ~= 4
error('normt_rnd: wrong # of input arguments');
end;

std = sqrt(sigma2);
% Calculate bounds on probabilities
  lowerProb = Phi((left-mu)./std);
  upperProb = Phi((right-mu)./std);
% Draw uniform from within (lowerProb,upperProb)
  u = lowerProb+(upperProb-lowerProb).*rand(size(mu));
% Find needed quantiles
```

```
  result = mu + Phiinv(u).*std;

function val=Phiinv(x)
% Computes the standard normal quantile function of the vector x, 0<x<1.

val=sqrt(2)*erfinv(2*x-1);

function y = Phi(x)
% Phi computes the standard normal distribution function value at x

y = .5*(1+erf(x/sqrt(2)));
```

References

Aarons L. (1999). Software for population pharmacokinetics and pharmacodynamics. *Clinical Pharmacokinetics* **36**, 255–264.

Atkins J.E. and Cheng C.S. (1999). Optimal regression designs in the presence of random block effects. *J. Statist. Plann. Inference* **77**, 321–335.

Atkinson A.C. and Bogacka B. (2002). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems* **61**, 17–33.

Atkinson A.C. and Cox D.R. (1974). Planning experiments for discriminating between models. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 321–348. With discussion by H. P. Wynn, D. M. Titterington, P. J. Laycock, D. V. Lindley, D. H. Hill, Agnes M. Herzberg, P. A. Tukey, A. O'Hagan, V. V. Fedorov, J. Dickey, J. Kiefer and C. A. B. Smith.

Atkinson A.C., Demetrio C.G.B., and Zocchi S.S. (1995). Optimum dose levels when males and females differ in response. *Applied Statistics* **44**, 213–226.

Atkinson A.C. and Donev A.N. (1992). *Optimum Experimental Designs*. Oxford University Press, Oxford.

Atkinson A.C. and Fedorov V.V. (1975a). The design of experiments for discriminating between two rival models. *Biometrika* **62**, 57–70.

Atkinson A.C. and Fedorov V.V. (1975b). Optimal design: experiments for discriminating between several models. *Biometrika* **62**, 289–303.

Atkinson A.C. and Haines L.M. (1996). Designs for nonlinear and generalized linear models. In *Design and analysis of experiments*, vol. 13 of *Handbook of Statistics*, pp. 437–475, North-Holland, Amsterdam.

Barone J.A., Koh J., Bierman R.H., Colaizzi J.L., Swanson K., Gaffar M., Moskovitz B.L., Mechlinski W., and Van de Velde V. (1993). Food interaction and steady-state pharmacokinetics of itraconazole capsules in healthy male volunteers. *Antimicrobial Agents and Chemotherapy* **37**, 778–784.

Barone J.A., Moskovitz B.L., Guarnieri J., Hassel A., Colaizzi J.L., Bierman R.H., and Jessen L. (1998). Food interactions and steady-state pharmacokinetics of itraconazole oral solution in healthy volunteers. *Pharmacotherapy* **18**, 295–301.

Bates D.M. and Watts D.G. (1988). *Nonlinear regression analysis and its applications*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York.

Berger M.P.F. and Tan F.E.S. (2004). Robust designs for linear mixed effects models. *Applied Statistics* **53**, 569–581.

Botev Z. and Kroese D.P. (2004). Global likelihood optimization via the cross-entropy method with an application to mixture models. In R.G. Ingalls, M.D. Rosetti, J.S. Smith, and B.A. Peters, eds., *Proceedings of the 2004 Winter Simulation Conference*.

Box G.E.P. and Lucas H.L. (1959). Design of experiments in non-linear situations. *Biometrika* **46**, 77–90.

Chaloner K. and Larntz K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference* **21**, 191–208.

Chaloner K. and Verdinelli I. (1995). Bayesian experimental design: a review. *Statist. Sci.* **10**, 273–304.

Cheng C.S. (1995). Optimal regression designs under random block-effects models. *Statist. Sinica* **5**, 485–497.

Clyde M.A. (2001). *International Encyclopedia of the Social & Behavioral Sciences*, chap. Experimental Design: Bayesian Designs, pp. 5075–5081. Elsevier Science Ltd, Oxford.

Cook R.D. and Wong W.K. (1994). On the equivalence of constrained and compound optimal designs. *J. Amer. Statist. Assoc.* **89**, 687–692.

Corana A., Marchesi M., Martini C., and Ridella S. (1987). Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Transactions on Mathematical Software* **13**, 262–280.

D'Argenio D.Z. (1981). Optimal sampling times for pharmacokinetic experiments. *Journal of Pharmacokinetics and Biopharmaceutics* **9**, 739–756.

Draper N.R. and Hunter W.G. (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika* **53**, 525–533.

Duffull S.B., Waterhouse T.H., and Eccleston J.A. (2005). Consideration of designs for population pharmacokinetic studies. Manuscript submitted for publication.

Finkelstein E. and Hall K. (1979). Aminoglycosides clearance in patients with cystic fibrosis. *Journal of Pediatrics* **94**, 163–164.

Gander W. and Gautschi W. (2000). Adaptive quadrature—revisited. *BIT* **40**, 84–101.

Goffe W.L., Ferrier G.D., and Rogers J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics* **60**, 65–99.

Green B. and Duffull S.B. (2003). Prospective evaluation of a d-optimal designed population pharmacokinetic study. *Journal of Pharmacokinetics and Pharmacodynamics* **30**, 145–161.

Hardin T.C., Graybill J.R., Fetchick R.J., Woestenborghs R., Rinaldi M.G., and Kuhn J. (1988). Pharmacokinetics of itraconazole following oral administration to normal volunteers. *Antimicrobial Agents and Chemotherapy* **32**, 1310–1313.

Hedman A., Adan-Abdi Y., Alvan G., Strandvik B., and Arvidsson A. (1988). Influence of the glomerular filtration rate on renal clearance of ceftazidime in cystic fibrosis. *Clinical Pharmacokinetics* **15**, 57–65.

Heykants J., Van Peer A., Van de Velde V., Van Rooy P., Meuldermans W., and Lavrijsen K. (1989). The clinical pharmacokinetics of itraconazole: an overview. *Mycoses* **32**, 67–87.

Hostetler J.S., Heykants J., Clemons K., Woestenborghs R., Hanson L.H., and Stevens D.A. (1993). Discrepancies in bioassay and chromatography determination explained by metabolism of itraconazole to hydroxyitraconazole: Studies of interpatient variations in concentrations. *Antimicrobial Agents and Chemotherapy* **37**, 2224–2227.

Janssen Pharmaceutica Products, L.P. (2002a). Sporanox® (itraconazole) injection. URL `http://www.fda.gov/cder/foi/label/2002/20966s6lbl.pdf`.

Janssen Pharmaceutica Products, L.P. (2002b). Sporanox® (itraconazole) oral solution. URL `http://www.fda.gov/cder/foi/label/2002/20657s7lbl.pdf`.

Jones B. and Wang J. (1998). Optimal cross-over designs for nonlinear mixed models using an approximate information matrix. *Journal of Combinatorics, Information & System Sciences* **23**, 379–391. J. N. Srivastava: felicitation volume.

Jones B. and Wang J. (1999). Constructing optimal designs for fitting pharmacokinetic models. *Statistics and Computing* **9**, 209–218.

Jones B., Wang J., Jarvis P., and Byrom W. (1999). Design of cross-over trials for pharmacokinetic studies. *Journal of Statistical Planning and Inference* **78**, 307–316.

Jusko W.J., Mosovich L.L., Gerbracht L.M., Matter M.E., and Yaffe S.J. (1975). Enhanced renal excretion of dicloxacillin in patients with cystic fibrosis. *Pediatrics* **56**, 1038–1044.

Kempton R.A., Ferris S.J., and David O. (2001). Optimal change-over designs when carry-over effects are proportional to direct effects of treatments. *Biometrika* **88**, 391–399.

Kiefer J. and Wolfowitz J. (1959). Optimum designs in regression problems. *Annals of Mathematical Statistics* **30**, 271–294.

Koks C., Sparidans R., Lucassen G., Crommentuyn K., and Beijnen J. (2002). Selective high performance liquid chromatography assay for itraconazole and hydroxyitraconazole in

plasma from human immunodeficiency virus-infected patients. *Journal of Chromatography B* **767**, 103–110.

Koks C.H.W., Huitema A.D.R., Kroon E.D.M., Chuenyam T., Sparidans R.W., Lange J.M.A., and Beijnen J.H. (2003). Population pharmacokinetics of itraconazole in thai hiv-1-infected persons. *Ther. Drug Monit.* **25**, 229–233.

Kroese D.P., Porotsky S., and Rubinstein R. (2005). The cross-entropy method for continuous multi-extremal optimization. Submitted.

Longford N.T. (1994). Logistic regression with random coefficients. *Computational Statistics & Data Analysis* **17**, 1–15.

McCullagh P. and Nelder J.A. (1989). *Generalized Linear Models*. Chapman and Hall, New York, 2nd ed.

Mentré F., Dubruc C., and Thénot J.P. (2001). Population pharmacokinetic analysis and optimization of the experimental design for mizolastine solution in children. *Journal of Pharmacokinetics and Pharmacodynamics* **28**, 299–319.

Mentré F., Mallet A., and Baccar D. (1997). Optimal design in random-effects regression models. *Biometrika* **84**, 429–442.

Moss R.B. (2002). Allergic bronchopulmonary aspergillosis. *Clinical Reviews in Allergy & Immunology* **23**, 87–104.

Müller W.G. and Ponce de Leon A.C.M. (1996). Discrimination between two binary data models: sequentially designed experiments. *Journal of Statistical Computation and Simulation* **55**, 87–100.

Myers R.H., Montgomery D.C., and Vining G.G. (2002). *Generalized linear models*. Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], New York. With applications in engineering and the sciences.

Nestorov I., Graham G., Duffull S., Aarons L., Fuseau E., and Coates P. (2001). Modeling and simulation for clinical trial design involving a categorical response: a phase II case study with naraptriptan. *Pharmaceutical Research* **18**, 1210–1219.

Ponce de Leon A.C. and Atkinson A.C. (1992). The design of experiments to discriminate between two rival generalized linear models. In L. Fahrmeir, B. Francis, R. Gilchrist, and G. Tutz, eds., *Advances in GLIM and Statistical Modelling*, vol. 78 of *Lecture Notes in Statistics*, pp. 159–164, Springer-Verlag, New York.

Press W.H., Teukolsky S.A., Vetterling W.T., and Flannery B.P. (2002). *Numerical recipes in C++*. Cambridge University Press, Cambridge. The art of scientific computing, Second edition, updated for C++.

Pukelsheim F. (1993). *Optimal design of experiments*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

Reed M.D. (1997). The pathophysiology and treatment of cystic fibrosis. *Journal of Pediatric Pharmacy Practice* **2**, 285–308.

Retout S. and Mentré F. (2003). Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics* **13**, 209–227.

Retout S., Mentré F., and Bruno R. (2002). Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Statistics in Medicine* **21**, 2623–2639.

Robinson L.D. and Jewell N.P. (1991). Some surprising results about covariate adjustment in logistic regression models. *International Statistics Review* **58**, 227–240.

Rubinstein R.Y. and Kroese D.P. (2004). *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*. Springer-Verlag.

Senn S. and Ezzet F. (1999). Clinical cross-over trials in phase i. *Statistical Methods in Medical Research* **8**, 263–278.

Sheiner L.B. and Beal S.L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. i. michaelis-menten model: routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics* **8**, 553–571.

Silvey S.D. (1980). *Optimal design*. Chapman & Hall, London. An introduction to the theory for parameter estimation, Monographs on Applied Probability and Statistics.

Skov M., Hoiby N., and Koch C. (2002). Itraconazole treatment of allergic bronchopulmonary aspergillosis in patients with cystic fibrosis. *Allergy* **57**, 723–8.

Smith K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* **12**, 1–85.

Spino M. (1991). Pharmacokinetics of drugs in cystic fibrosis. *Clinical Reviews in Allergy* **9**, 169–210.

Stevens D.A., Lee J.Y., Schwartz H.J., Jerome D., and Catanzaro A. (2000). A randomized trial of itraconazole in allergic bronchopulmonary aspergillosis. *The New England Journal of Medicine* **342**, 756–762.

Stigler S.M. (1971). Optimal experimental design for polynomial regression. *Journal of the American Statistical Association* **66**, 311–318.

Studden W.J. (1982). Some robust-type D-optimal designs in polynomial regression. *J. Amer. Statist. Assoc.* **77**, 916–921.

Uciński D. and Bogacka B. (2002). Construction of T-optimum designs for multiresponse dynamic models. In *COMPSTAT 2002 (Berlin)*, pp. 267–272, Physica, Heidelberg.

Van Cutsem J. (1989). The in-vitro antifungal spectrum of itraconazole. *Mycoses* **32**, 7–13.

Van de Velde V., Van Peer A., Heykants J., Woestenborghs R., Van Rooy P., De Beule K., and Cauwenbergh G. (1996). Effect of food on the pharmacokinetics of a new hydroxypropyl β-cyclodextrin formulation of itraconazole. *Pharmacotherapy* **16**, 424–428.

Van Peer A., Woestenborghs R., Heykants J., Gasparini R., and Gauwenbergh G. (1989). The effects of food and dose on the oral systemic availability of itraconazole in healthy subjects. *European Journal of Clinical Pharmacology* **36**, 423–426.

Wald A. (1943). On the efficient design of statistical investigations. *Annals of Mathematical Statistics* **14**, 134–140.

Waterhouse T.H., Eccleston J.A., and Duffull S.B. (2004a). On optimal design for discrimination and estimation. In J. Antoch, ed., *COMPSTAT 2004 - Proceedings In Computational Statistics: 16th Symposium Held In Prague, Czech Republic, 2004*, pp. 1963–1970, Physica-Verlag.

Waterhouse T.H., Eccleston J.A., and Duffull S.B. (2004b). On optimal design for discrimination and estimation. *Research Report 106*, Centre for Statistics, The University of Queensland.

Waterhouse T.H., Redmann S., Duffull S.B., and Eccleston J.A. (2005). Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients. *Journal of Pharmacokinetics and Pharmacodynamics* (in press).

Yano I., Beal S.L., and Sheiner L.B. (2001). The need for mixed-effects modeling with population dichotomus data. *Journal of Pharmacokinetics and Pharmacodynamics* **28**, 389–412.